Department of Education and Science
Department of Education for Northern Ireland
Welsh Office

Assessment of Performance Unit

APU

Science in Schools

Age 15: Report No. 1

by

Rosalind Driver
Richard Gott
Sandra Johnson
Christopher Worsley
Fiona Wylie

Report on the 1980 survey of science performance of pupils aged 15 to the Department of Education and Science, the Department of Education for Northern Ireland and the Welsh Office.

Centre for Studies in Science Education, University of Leeds

Chelsea College, Centre for Science and Mathematics Education, University of London.

REFERENCE ONLY
THIS BOOK MAY NOT BE
TAKEN OUT OF THE LIBRARY

London
Her Majesty's Stationery Office

SOUTHAMPTON UNIVERSITY LIBRARY

© *Crown copyright 1982*
First published 1982

The research teams

The teams are based at two centres, at the University of Leeds and at Chelsea College, University of London. The work of monitoring at three ages (11, 13 and 15) in science is divided between these centres so that the responsibility for monitoring at the lower two age groups rests with Chelsea and the responsibility for monitoring at age 15 and for the data processing and analysis rests with Leeds.

The science monitoring teams:

*At Leeds**
David Layton, Director**
Rosalind Driver, Deputy Director
Roger Hartley, Deputy Director
Richard Gott
Jennifer Head (Statistician)
Sandra Johnson (Statistician)
Brian Maher (Programmer)
Christopher Worsley
Fiona Wylie

Secretarial staff
Glynis Wilkinson
Helen Johnson
Elizabeth Lodge

At Chelsea
Paul Black, Director
Wynne Harlen, Deputy Director
Patricia Murphy
Tony Orgee
David Palacio
Terry Russell
Beta Schofield

Past members of the teams
Brenda Denvir (Jan to Aug 1981)
Cynthia Millband (1980 – 1981)
Ardrie VanderWaal (1979 – 1980)

*Membership of Leeds team at the time the survey report was prepared
**Dennis Child, Director from May 1982

Much of the work in this report is the product of planning and development in which all members of the teams (listed above) and the Steering Group on Science (listed in Appendix 10) have had a part.

In producing the report, Rosalind Driver drafted Chapters 1, 6, 7 and 10, and also acted as overall report editor. Christopher Worsley produced all of the individual question results quoted throughout, and drafted Chapters 4 and 5. Sandra Johnson analysed the school questionnaire, produced the sub-category test results and drafted Chapters 2 and 9. Chapter 3 was drafted by Richard Gott and Chapter 8 by Fiona Wylie.

ISBN 0 11 270393 3

Acknowledgements

The monitoring team would like to thank the many people who contributed to the various phases of work involved in conducting the survey and producing this report.

The monitoring framework and test questions used were the result of a number of years of development work co-operatively undertaken by the teams at Chelsea College and at the University of Leeds in collaboration with the Science Steering Group.

Liaison groups of science teachers and members of staff at the Science Education Centre at Leeds worked regularly with the team in developing and criticising test questions.

Mrs B A Bloomfield and her staff at the Monitoring Services Unit were responsible for the drawing of the sample and subsequent administrative arrangements with the survey schools.

Science Advisers in the Local Education Authorities helped by nominating teachers as practical testers and arranging for their release from normal teaching duties. In Northern Ireland, Mr I W Milligan HMI made arrangements for the recruitment and training of testers. Dr N B Evans HMI recruited the practical testers in Wales and personally undertook the translation of the test materials into Welsh.

Philip Harris Ltd. supplied the apparatus for the practical testing. The care and ingenuity of members of their staff have been greatly appreciated.

The administration of the practical tests, a demanding part of the exercise, was undertaken most competently by science teachers from England, Wales and Northern Ireland, as was the task of marking the written tests.

The conduct of the survey itself depended on the goodwill and co-operation of pupils and staff in the survey schools. We are particularly grateful to the teachers, who undertook the various time-consuming tasks involved in administering the tests, and without whose help this survey would not have been possible.

Contents

Page

Preface xi

1. **Introduction**
   - 1.1 Background to the survey 1
   - 1.2 The assessment framework 1
   - 1.3 The 1980 survey 3
   - 1.4 Features of the report 4

2. **Science provision in the survey schools**
   - 2.1 Introduction 6
   - 2.2 The schools and their resources 8
   - 2.3 The pupils' science course choices 16
   - 2.4 Summary 22

3. **Using symbolic representations**
   - 3.1 Introduction 24
   - 3.2 Reading information from graphs, charts and tables: sub-category 1α 24
   - 3.3 Expressing information as graphs, charts and tables: sub-category 1β 26
   - 3.4 Summary of scores: sub-categories 1α and 1β 28
   - 3.5 Description of performance levels on selected question types 28

4. **Using apparatus and measuring instruments**
   - 4.1 Introduction 43
   - 4.2 Administration 44
   - 4.3 Using laboratory instruments: sub-category 2α 45
   - 4.4 Estimating: sub-category 2β 51
   - 4.5 Following instructions: sub-category 2γ 59
   - 4.6 Summary of Category 2 64

5. **Using observation**
   - 5.1 Introduction 65
   - 5.2 Administration 66
   - 5.3 Using a branching key: sub-category 3α 66
   - 5.4 Observing similarities and differences: sub-category 3β 69
   - 5.5 Interpreting observations: sub-category 3γ 75
   - 5.6 Summary of Category 3 80

6. Interpretation of data

6.1 Introduction 81
6.2 Describing and using patterns in information: sub-category 4α 81
6.3 Judging the applicability of a given generalisation: sub-category 4β 90
6.4 Distinguishing degrees of inference: sub-category 4γ 92
6.5 Summary of results for data interpretation 93

7. Application of science concepts

7.1 The nature of the category 94
7.2 The concept base 94
7.3 Description of the questions and tests 95
7.4 Applying biological concepts: sub-category $4\delta_{AB}$ 95
7.5 Applying physics concepts: sub-category $4\delta_{CD}$ 105
7.6 Applying chemistry concepts: sub-category $4\delta_{EF}$ 117
7.7 Generating alternative hypotheses: sub-category 4ε 125
7.8 Generating comments on the interpretation of question scores 128

8. Design of investigations

8.1 Introduction 129
8.2 Assessing testable statements: sub-category 5α 129
8.3 Assessing experimental procedures: sub-category 5β 134
8.4 Summary of performance on 5α and 5β 143
8.5 Devising and describing investigations: sub-category 5γ 144
8.6 Discussion of performance on 'design of investigations' 151

9. Pupils' science performances

9.1 Introduction 152
9.2 Pupil performance in relation to general school characteristics 154
9.3 Pupil performance in relation to science-specific school characteristics 159
9.4 Sex and curriculum background 163
9.5 Summary 166

10. Interpretation and discussion of the findings

10.1 Levels of performance 168
10.2 Science curriculum groups and science performance 169
10.3 Sex differences in performance 171
10.4 Performance in relation to school characteristics 172
10.5 Monitoring in science: future plans 173
10.6 Implications for science education: matters for discussion 174

Appendices

1 The assessment framework
Description of the categories, sub-categories and component types of questions 180

The list of concepts and knowledge 192
The list of apparatus and measuring instruments 203
2 Development and selection of the science questions 206
3 The organisation of the group practical tests 208
4 Administration and marking of written tests 212
5 Notes on the sampling of schools and pupils 214
6 The questionnaire 218
7 Notes on the survey analysis 222
8 Standard errors, confidence limits and statistical significance 225
9 Administrators of practical tests 227
10 Members of the APU Steering Group on Science 228
11 Monitoring Services Unit 229
12 Members of the APU Consultative Committee 230
13 A Note on the APU 231

Tables

Page

1.1 Framework of categories and sub-categories 2
2.1 Classification of pupil sample into five examination-entry groups 17
2.2 The relative popularity of science course combinations 17
2.3 The relative popularity of specific science subjects (subjects taken in any course combination) 18
3.1 Drawing: data averaged for six questions 37
3.2 Data for question where one point did not lie on the smooth curve 37
4.1 Reading scales on laboratory instruments 48
4.2 Table of performance by pupils on estimating quantities 53
4.3 Breakdown of performance on the question 'wiring a plug' 60
4.4 Breakdown of pupils' performances when lighting a bunsen burner 62
4.5 Breakdown of pupils' performances when heating a solid 62
4.6 Breakdown of pupils' performances when using non-contamination techniques 63
4.7 Breakdown of pupils' performances when testing for a gas 63
5.1 Score distributions for questions involving drawing 74
5.2 Percentage of pupils in each category 77
8.1 Methods suggested by pupils (ping pong) 145
8.2 Features of 'rebound height' method 146
8.3 Combinations on 'rebound height' method 147
8.4 Features of 'number of bounces' method 148
8.5 Combinations on 'number of bounces' method 148
8.6 Pupils' suggested methods (fur kettle) 149
8.7 Features of 'changing the element' method 149
8.8 Combinations on 'changing the element' method 150
8.9 Features of 'changing the water' method 150
8.10 Combinations on 'changing the water' method 151
A5.1 School participation 214
A5.2 Pupil and school samples at the analysis stage 217
A7.1 Summary of some of the survey results 224

Figures

Page

2.1 The five regions 7

2.2 Pupil – teacher ratio (% schools within region) 9

2.3 Pupil – teacher ratio (% schools within type in England) 9

2.4 Pupil – teacher ratio (% schools within location in England) 10

2.5 Adequate laboratory accommodation (% schools within region) 11

2.6 Adequate laboratory accommodation (% schools within type in England) 11

2.7 Adequate level of technician support (% schools within region) 12

2.8 Adequate level of technician support (% schools within type in England) 13

2.9 Average annual expenditure per pupil on all teaching resources (% schools within region) 14

2.10 Average annual expenditure per pupil on all teaching resources (% schools within type in England) 14

2.11 Average annual science expenditure: £ per pupil on roll (% schools within region) 15

2.12 Average annual science expenditure: £ per pupil on roll (% schools within type in England) 15

2.13 Sex differences in science course choices 19

2.14 Proportions of boys and girls studying each science course in any combination 19

2.15 Examination-entry grade distributions for different science course combinations 20

2.16 Regional distributions across examination-entry grades 21

4.1 Pupils' estimates of length 54

4.2 Pupils' estimates of time 54

4.3 Pupils' estimates of circular area 55

4.4 Pupils' estimates of magnification 55

4.5 Pupils' estimates of a large mass 56

4.6 Pupils' estimates of a small mass 56

4.7 Pupils' estimates of force 57

4.8 Pupils' estimates of volume 57

4.9 Pupils' estimates of candle flame temperature 58

4.10 Proportion of pupils estimating in the range 1 – 208°C 58

9.1 Overall sub-category performance levels (all 15 year old pupils in England, Wales and Northern Ireland) 154

9.2 Sub-category performances by school location (15 year old pupils in England) 156

9.3 Regional patterns of sub-category performance 157

9.4 Percentage free school meals uptake and sub-category performance (15 year old pupils in comprehensive schools in England) 158

9.5 Laboratory accommodation and sub-category performance (15 year old pupils in comprehensive schools in England) 160

9.6 Technician support and sub-category performances (15 year old pupils in comprehensive schools in England) 161

9.7 Schools' average annual science expenditure and sub-category performances (15 year old pupils in comprehensive schools in England) 162

9.8 Sex differences in sub-category performance (15 year old pupils in England, Wales and Northern Ireland) 164

9.9 Curriculum group differences in sub-category performance: generalised pattern 165

A5.1 The regional composition of the school sample and population 215

A5.2 Composition of English school sample and population in terms of type of school 215

Preface

This is the first report by the Assessment of Performance Unit on 15 year olds' performance in science. It covers the survey carried out in November 1980.

The report describes how a national sample of 15 year olds drawn from all levels of ability coped with a wide range of questions covering scientific processes and knowledge. It also contains information about curriculum organisation and the physical, financial and human resources for science activities of the schools surveyed. The authors invite thought and discussion on the interpretation of the results revealed by the survey, but do not draw conclusions from those results. It is for those to whom this report is addressed—educationists, parents, and those concerned with the provision of resources centrally and locally—to consider how far the standards of performance reflected in the report are acceptable.

To date, the APU has published one report on the scientific performance of 11 year olds and one on that of 13 year olds. The APU will be undertaking further annual surveys which, together, should provide a clear picture of what pupils can achieve in science at the ages of 11, 13 and 15.

The report is not only intended for science educators but for all those with a professional interest. We hope that readers will find the report interesting in itself and will also be encouraged to relate it to APU reports of surveys in other areas of the curriculum.

Jean Dawson A G Clegg

Jean Dawson Arthur Clegg
Joint heads of the Assessment of Performance Unit

1 Introduction

1.1 Background to the survey

This survey forms part of a national monitoring programme set up by the Assessment of Performance Unit of the Department of Education and Science. The tasks outlined for the programme included the development of new assessment techniques, the conduct of national surveys and the use of these to identify significant differences in achievement related to the circumstances in which children learn (DES 1978a).

The design of the framework for the assessment in science was the responsibility of the two monitoring teams, (which are based at the University of Leeds and at Chelsea College, University of London), working within a policy established by the Science Steering Group set up by the Assessment of Performance Unit.

The team at Chelsea College is responsible for the science surveys at ages 11 and 13. The first surveys at these ages were conducted in May and June 1980 and have already been reported (DES 1981a, 1982a). The monitoring team at the University of Leeds is responsible for the surveys at age 15 and for the analysis of the results at all three ages.

This report presents the results of the first survey of scientific performance of 15 year old pupils. The survey was conducted in November 1980 and involved 16,301 pupils in more than 500 randomly selected schools in England, Wales and Northern Ireland. The purpose of the survey was not to assess individuals or schools but to give a picture of the performance of pupils in this age group; the result is a profile report, not of individuals but of the population of 15 year old pupils as a whole, on a range of types of science activities.

1.2 The assessment framework

One of the first tasks for the science teams after their appointment in 1977 was to define and agree what was to be assessed in the monitoring programme. After a period of consultation and experimentation with test materials, an outline for the assessment framework was agreed. This was published in the Science Progress Report (DES 1978b). Six main categories of activity were identified which encompass both general methods of scientific enquiry and the application of a body of knowledge and concepts relating to the natural world. Most of these categories were further divided producing a total of 19 sub-categories which are tested in this survey. This framework of categories and sub-categories is outlined in Table 1.1.

The first two categories of activity concern specific skills. In science, information is represented in a variety of ways. Graphs, charts and diagrams

Table 1.1 *The framework of categories and sub-categories used in the assessment of performance in science at age 15*

Categories		Major sub-categories		Minor sub-categories
1. Using symbolic representations *(Written tests)*	α	Reading information from graphs, tables and charts	γ	*Using scientific symbols and conventions*
	β	Expressing information as graphs, tables and charts		
2. Using apparatus and measuring instruments *(Practical tests)*	α	Using measuring instruments	β	*Estimating quantities*
			γ	*Following instructions for practical work*
3. Using observations *(Practical tests)*	β	Observing similarities and differences	α	*Using an identification key*
			γ	*Interpreting observations*
4. Interpretation and application				
I Not dependent on taught science concepts *(Written tests)*	α	Describing and using patterns in information	β	*Judging the applicability of a given generalisation*
			γ	*Distinguishing degrees of inference*
II Dependent on taught science concepts *(Written tests)*	δ_{AB}	Applying biology concepts	ε	*Generating alternative hypotheses*
	δ_{CD}	Applying physics concepts		
	δ_{EF}	Applying chemistry concepts		
5. Design of investigations *(Written tests)*	α	Assessing testable statements	γ	*Devising and describing investigations*
	β	Assessing experimental procedures		
6. Performing investigations *(Practical tests)*		Questions in this category are to be introduced in the 1981 survey		

For major sub-categories an overall sub-category score is computed.
For minor sub-categories no overall score is computed; the sub-category is represented by questions reported in full.

are used to communicate information in a clear and concise manner. Category 1, 'Using symbolic representations', is concerned with the transformation of data from one form to another: reading information off graphs, charts and tables and putting data into graphical or tabular form. It also includes knowledge of conventional notation specific to certain areas of science, for example circuit diagrams or chemical symbols. Experimental work in science depends on the appropriate use of a range of apparatus and measuring instruments, as well as an appreciation of the magnitude of physical quantities. Category 2 is concerned with such practical skills.

Skills associated with making and interpreting observations, such as discerning similarities and differences between groups of objects, grouping or classifying objects based on observable properties, and recording a sequence of events are the focus of Category 3 'Using observation'. Once data has been collected in a systematic way it can be inspected for patterns, trends or apparent relationships. The first part of Category 4 'Interpretation and application' is concerned with these data interpretation skills.

It is generally accepted that one of the main aims of science education in secondary schools is to introduce pupils to a coherent system of scientific knowledge and concepts, and to give pupils an understanding of ideas which will both form a foundation for future studies in science and be of use in their everyday lives. The use of a range of scientific concepts in making predictions and giving explanations is the concern of the second part of Category 4. The ability to use and apply science concepts is reported separately for biology, physics and chemistry.

Category 5 concerns aspects of experimental design. It involves the framing of questions so they are in a testable form, as well as the control of appropriate variables, the sequencing of procedures and the minimising of sources of error.

It was decided that performing practical investigations was different in significant ways from designing them and therefore a separate category, Category 6, is devoted to the assessment of practical investigation skills. Tests relating to this category were not used in the 1980 survey but will be introduced in 1981.

In this assessment framework it is not assumed that activities in science such as making observations, interpreting data or designing experiments are necessarily carried out independently of a framework of scientific knowledge. A list of the knowledge and concepts used for the purpose of the assessment was agreed through consultation with science teachers. The way the list was developed is described in Chapter 7 and the list is given in Appendix 1.

The methods of testing have been selected to be appropriate to the activity being tested. Written tests are used for Categories 1, 4 and 5. Categories, 2, 3 and 6 are assessed by practical tests.

1.3 The 1980 survey

The questions used in the survey were drawn from the bank of questions which had been developed and trialled by the team. (For details of the question development procedures, readers are referred to Appendix 2.)

A total of 20 different written test packages and 5 practical tests were used in the survey, each designed to take about one hour of testing time. Each type of test was administered to a different sub-sample of pupils so that no pupil or school saw the complete range of test packages. The majority of the pupils selected for testing were aged 15 when the tests were administered in November

1980 (for convenience, therefore, we refer to the pupils tested in the survey as 15 year old pupils).

The Monitoring Services Unit at the National Foundation for Educational Research was responsible for drawing the samples of pupils, corresponding with the schools, and distributing and collecting the tests. It also planned the itineraries of the practical testers and made administrative arrangements with the sample schools.

Individual pupils taking part in the survey remain anonymous and the schools and LEAs involved are not identified.

1.4 Features of the report

Chapter 2 describes the sampling procedure and gives information about the pupils and the schools involved in the survey. In order to aid the interpretation of the results, information about the provision for science teaching was collected through a questionnaire sent to each participating school. The findings from the questionnaire are also described in this chapter. (The questionnaire itself is reproduced in Appendix 6).

Altogether over 350 questions grouped into 25 tests were used in the survey; each question being given to over 700 pupils. The data from these questions have been analysed and reported on in a number of ways. In each of ten of the sub-categories in this survey, the questions were considered to be assessing the same skill or ability, and an overall sub-category score was produced. These sub-categories are designated "major sub-categories" in Table 1.1. The resulting sub-category scores were analysed and reported by various background variables.

Within these sub-categories smaller groups of questions have been used to obtain descriptive information on particular skills giving a more detailed analysis of levels of performance and types of error. This has been attempted on a small scale in this report. In future years it is planned to extend this type of reporting.

In nine sub-categories designated as 'minor' in Table 1.1, not enough questions were used in the survey to produce an overall score. In these cases individual questions were selected for reporting in full either for their inherent interest or as exemplars of questions of a type.

Chapters 3 – 8 are devoted to a description of pupils' performances on the various categories of activity. In each of the chapters, examples of questions are given together with information about pupils' levels of performance. These example questions, numbered sequentially throughout each chapter, are displayed together with results obtained and information needed to interpret them.

Readers will notice, when referring to these examples, that questions varied in the maximum number of marks allocated. In order to give equal weighting to each question contributing to a sub-category score, the raw marks were mapped on to a scale with a maximum of either 3 or 1. (For further details of this readers are referred to Appendix 2). Where relevant, the way raw scores are mapped on to the scaled scores is indicated for questions contributing to a sub-category score. Information on each question in the report includes the mean score, expressed as a percentage, the total number of pupils to whom the question was given (n), and the percentage of pupils who did not attempt the question.

Overall performance profiles for the ten major sub-categories are given in Chapter 9. These are reported for various pupil sub-groups, (eg. for girls and boys, and for different regions of the country). Patterns relating levels of performance with levels of provision for science and with other school variables are also described.

In the last chapter the main findings of the survey are summarised and discussed. Some interpretations are suggested and an attempt is made to identify some issues for future consideration.

The purpose of the survey has not been to assess individual pupils or schools, but to report levels of performance in the population of 15 years olds as a whole. In contrast with the results for public examinations in which individual pupils are assessed in relation to one another, the results in this survey are reported in such a way as to indicate levels of performance on a range of science related activities in relation to the demands of the tasks set.

Each sub-category is assessing a different kind of activity. It is not, therefore, appropriate to make direct comparisons between the levels of performance on different sub-categories: each has to be judged independently taking into account the types of questions and mark schemes used.

It is recognised that the levels of performance on the range of science related activities assessed in this survey may be the result of experiences other than those obtained in science lessons. Other school subjects as well as pupils' out-of-school experiences may contribute to the pattern of results. In particular mathematics teaching is likely to contribute to performance on graphical skills and work in craft may contribute to pupils' skills of measurement and use of apparatus. Also pupils may gain experience in data interpretation in a range of subjects including geography, social studies and mathematics. For this reason, it is hoped that both science teachers and teachers of other subjects may find the results in this report to be of interest.

2 Science provision in the survey schools

2.1 Introduction

There can be great variety among 15 year olds both in the nature of the science they are studying and in the overall time spent on this area of the curriculum. Some of these pupils will not have studied any science at all during the previous eighteen months, while those who have chosen to continue their science education often have a wide range of science subjects available to them.

Some information about the nature and extent of the pupils' science education was collected by requesting each school to provide details of the science courses being studied by each of its sample pupils. In addition, every school which participated in the survey was asked to complete a questionnaire concerning its teaching, physical and financial resources for work in science. All but 19 of the 568 participating schools returned a completed questionnaire, and all but 10 schools provided the requested curriculum information.

Before presenting the results of this enquiry, it may be pertinent to note one or two features of the school sample which should be borne in mind when the resource findings are considered.

The pupil sample was selected by a two-stage procedure. First a random sample of schools was drawn from a list of all appropriate schools containing 15 year old pupils in England, Wales and Northern Ireland. Pupils were then chosen randomly from within the relevant age-band in each selected school.

Before selection, schools were first classified by region of the country, and, within regions, by type and size. Wales and Northern Ireland were each considered regions for this purpose, and three more regions were formed by dividing England into the geographical areas of North, Midlands and South as shown in Figure 2.1.

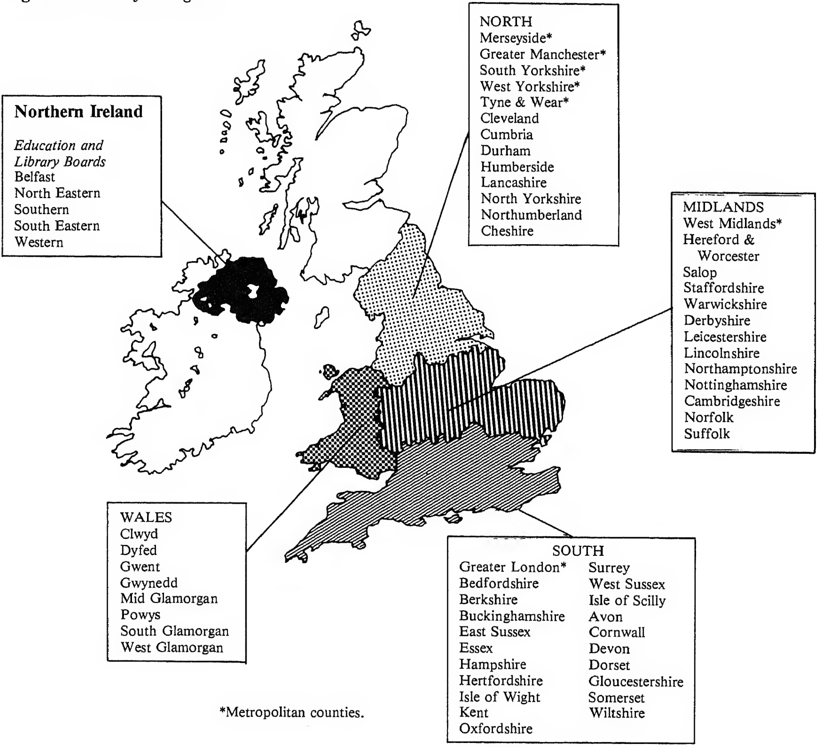
Schools were randomly selected from within these groups in numbers which appropriately reflected relative group sizes, except that Wales and Northern Ireland were deliberately over-sampled so that their pupil sample sizes would be large enough to produce regional estimates of performance of similar accuracy to those of the English regions.

In total, 16,301 pupils were included in the testing programme, these being drawn from 568 participating schools. (Further details of the sampling procedure and of the composition of the sample are to be found in Appendix 5).

An account is given in this chapter of *regional* patterns of resource provision and science courses studied, these figures being based on regional samples containing all types of school. Within England only, comparisons are made between the resource patterns of schools and the curriculum choices of pupils in metropolitan and non-metropolitan *locations.*

Proportionally more of the schools in the North are in metropolitan locations compared with those in the South and Midlands (a feature suggested in Figure 2.1). This imbalance in both the English school sample and its population may result in some overlap of regional and location findings.

Figure 2.1 *The five regions*



Northern Ireland

Education and Library Boards
Belfast
North Eastern
Southern
South Eastern
Western

NORTH
Merseyside*
Greater Manchester*
South Yorkshire*
West Yorkshire*
Tyne & Wear*
Cleveland
Cumbria
Durham
Humberside
Lancashire
North Yorkshire
Northumberland
Cheshire

MIDLANDS
West Midlands*
Hereford & Worcester
Salop
Staffordshire
Warwickshire
Derbyshire
Leicestershire
Lincolnshire
Northamptonshire
Nottinghamshire
Cambridgeshire
Norfolk
Suffolk

WALES
Clwyd
Dyfed
Gwent
Gwynedd
Mid Glamorgan
Powys
South Glamorgan
West Glamorgan

SOUTH
Greater London*
Bedfordshire
Berkshire
Buckinghamshire
East Sussex
Essex
Hampshire
Hertfordshire
Isle of Wight
Kent
Oxfordshire
Surrey
West Sussex
Isle of Scilly
Avon
Cornwall
Devon
Dorset
Gloucestershire
Somerset
Wiltshire

*Metropolitan counties.

Resource comparisons for different *types* of school, and subject choice comparisons for the pupils in these schools, also refer to England only. This is because the organisational types (eg. grammar, comprehensive, independent) which apply in England and Wales are not relevant in Northern Ireland (where instead a management system operates), while in Wales fully three-quarters of all secondary schools are comprehensive, and most of these are all-through comprehensives.

The types of school considered are three kinds of comprehensive distinguished by age range: 11 – 16 (includes 13 – 16), 13 – 18 and 11 – 18, modern and independent. There were too few grammar schools in the survey sample (20) for meaningful breakdowns to be given for this type of school, but occasionally comments are made about the resources of these schools where this seems appropriate.

The second regional imbalance within England which is relevant to note at this point is that the South region has proportionally more independent schools, and proportionally fewer comprehensive schools, than either the Midlands or North.

2.2 The schools and their resources

2.2.1 Introduction

The school questionnaire (which is reproduced as Appendix 6) collected information about some of the physical and financial resources which each survey school had available for science education. In addition, it was possible to gain information about such general school characteristics as pupil-teacher ratio and type of school from existing records.

With continuing comprehensive reorganisation in mind, one of the questions to each school was whether it had been affected by a restructuring of any kind which might explain unusual capitation figures. While very few of the sample schools in Wales and Northern Ireland claimed to be in a state of change, the proportion of affected schools in the English sample was large enough to justify comment.

About 15% of the English survey schools were undergoing, or had experienced during the previous five years, a change of one kind or another as a direct result of reorganisation along comprehensive lines. 45 schools were in the process of becoming comprehensive, just under one-third of these by merger with other schools. Six of the schools were new, another six were about to establish their sixth forms for the first time, and another half dozen were experiencing gradually rising or falling rolls as a result of reorganisation. Some schools (22 in all) were receiving new science laboratories and other equipment, usually in readiness for the anticipated emergence of their sixth forms.

This degree of change should be borne in mind when considering the resource findings which follow.

2.2.2 Teaching resources

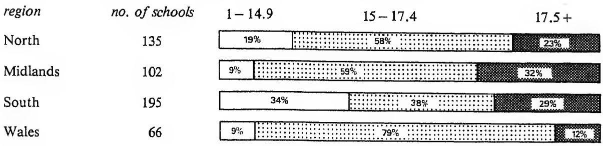
Overall pupil/teacher ratio statistics were available from official sources for all of the sample schools in England and Wales. In addition, each school was asked to indicate the number of teachers who were currently involved in teaching science in the school, and, of these, the number holding a science qualification. Information was also sought on laboratory availability and technician help.

Pupil/Teacher ratio

Acknowledging the variations known to exist in the way pupil/teacher ratio statistics are computed by different LEAs, and recognising that pupil/teacher ratios cannot be assumed to bear any direct relationship with sizes of teaching groups within schools, it was nevertheless felt that some information about this index would be of interest. In order to maintain comparability with previous APU survey reports, schools were classified into three ranges in terms of their pupil/teacher ratios: 1 – 14.9, 15 – 17.4 and 17.5 + .

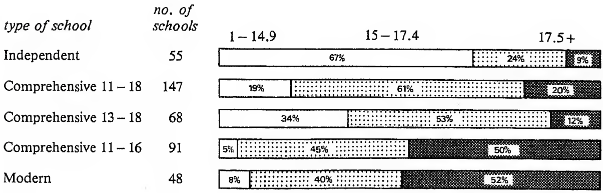
Figure 2.2 illustrates the regional differences which exist in the school sample for this variable, with proportionally more of the sample schools in the South of England having more favourable ratios than those in the other regions. (There are no pupil/teacher ratio statistics available for the schools in Northern Ireland.)

Figure 2.2 *Pupil/teacher ratio (% schools within region)*



If the different *types* of school in the English sample are considered, we see from Figure 2.3 that proportionally more of the 11 – 16 comprehensives and of the modern schools were in the group with the *least favourable* pupil/teacher ratios. The few grammar schools in the sample resembled the 11 – 18 comprehensives in having more favourable ratios. This pattern, distinguishing as it does between schools with and without sixth forms, is perhaps to be expected.

Figure 2.3 *Pupil/teacher ratio (% schools within type in England)*



The distribution for the independent schools is different again, with a relatively high proportion of these schools having the lowest pupil/teacher ratios.

Figure 2.4 *Pupil/teacher ratio (% schools within location in England)*

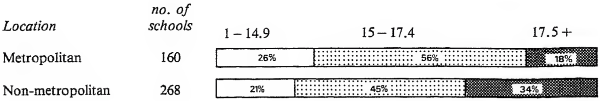


Figure 2.4 reveals some location difference with proportionally more of the non-metropolitan schools than metropolitan ones falling into the group with the least favourable ratios.

Qualified science teachers

Each survey school was asked for the number of teachers currently teaching science, and was also asked to indicate how many of these members of staff held a science qualification (science or applied science degree or teaching certificate with science as main subject).

Overall, about 95% of the teachers who were teaching some science in the survey schools held one of these science qualifications, with little regional variation in this proportion.

The modern schools and the 11 – 16 comprehensives in the survey sample had the lowest proportions of qualified science teachers, although this was nevertheless high at 91%. 95% of those teaching science in comprehensive schools with sixth forms had an appropriate qualification. In the independent schools (and, incidentally, in the grammar schools) virtually all of the teachers teaching science were qualified in the subject.

There was no school location difference in terms of their proportions of qualified science teachers.

Laboratories and technicians

Each school was asked to indicate the number of science laboratories it had available, and the number of technicians it employed to service these laboratories.

In order to explore laboratory provision in the schools, HMI's criterion of adequacy of laboratory accommodation was adopted (DES 1979). It was necessary to adapt this criterion to provide an index of adequate laboratory provision for those survey schools with ages of intake above 11. The HMI criterion is based on the assumption that pupils aged from 11 to 16 will study science for about a sixth of their curriculum time (ie. 6 periods in a 35 period week) and that their science work will be practically based. Allowing one period per day for general laboratory maintenance, there will be available in each laboratory about 30 hours of teaching time in a 35-period week or 35 hours in a 40-period week. An n-form entry school with y year-groups within the age-range 11 – 16 would then need 6ny/30 or 7ny/35 (ie. ny/5) laboratories in order to accommodate such a practically-based science teaching load.

Applying this criterion, a 5-form entry 11 – 16 comprehensive school would need 5 laboratories, while a 3-form entry 13 – 16 school would need about 2 laboratories to be considered adequately provided. HMI considered that schools with sixth forms would need one further laboratory over this basic number to meet their extra requirements. When evaluating the accommodation in their sample schools, HMI had available to them information about the particular circumstances obtaining in each, and the result of applying this general criterion was often modified for individual schools in the light of this information. It should be noted that similar circumstantial information was not available for the present survey schools.

Figure 2.5 *Adequate laboratory accommodation (% schools within region)*

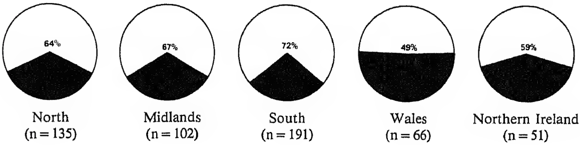
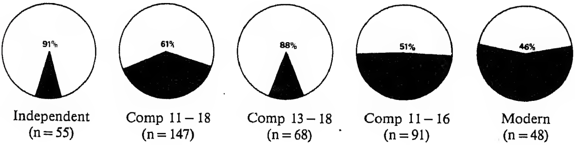


Figure 2.5 provides the regional picture of laboratory accommodation. About two-thirds of the survey schools in England had adequate numbers of laboratories according to the HMI criterion, with little variation from region to region. Proportionally fewer of the survey schools in Northern Ireland had sufficient numbers of laboratories for their needs, whilst in Wales only half of the survey schools had adequate laboratory accommodation.

Figure 2.6 *Adequate laboratory accommodation (% schools within type in England)*



About two-thirds of the comprehensive schools in England had adequate accommodation according to this criterion—the proportion previously found for the English comprehensive schools containing 15 year old pupils (the 'senior' comprehensives) in the school sample for the 1980 age 13 science survey (DES 1982a). The proportion varies, though, from one type of comprehensive school to another, as Figure 2.6 shows. The 13 – 18 schools had the highest levels of adequate laboratory accommodation, the 11 – 16 comprehensives the lowest. The independent schools were in general well-accommodated (as were the grammar schools in the sample).

There was little location difference in this respect, a slightly higher proportion of the survey schools in non-metropolitan locations in England having adequate laboratory accommodation compared with those in metropolitan locations (64% and 69% respectively).

The arrangements for technician support vary widely from school to school. Technicians can be employed full-time or part-time. Some work for the complete year, others only during term time. In order to obtain some indication of technician support, each school was asked to record for each of its science technicians the number of weeks worked per year, and the number of hours worked per week in the science department.

The questionnaire revealed that 85% of all the technicians worked for 40 weeks or more per year, and two-thirds of these were full-time.

Again, the HMI secondary school survey report (DES 1979) provided a useful criterion for judging the adequacy of a school's technician support, schools with three or fewer laboratories per full-time technician being considered by HMI to be adequately supported. Given the variety of employment conditions pertaining in the schools, it was necessary to determine the number of full-time equivalent technicians working in each school before computing this laboratory-technician ratio, ('full-time' being defined as a 35-hour week for 40 weeks or more per year).

Figure 2.7 shows the proportions of schools in the different regions with adequate technician support, based on two kinds of laboratory-technician ratio: in one case the ratio takes no account of the conditions of employment of the technicians concerned, the other is in terms of full-time equivalents. The discrepancies between the results for the two kinds of ratio reflect the extent to which the different regions employ technicians on a part-time basis—the larger the difference the more prevalent is part-time employment in that region.

Figure 2.7 *Adequate level of technician support (% schools within region)*

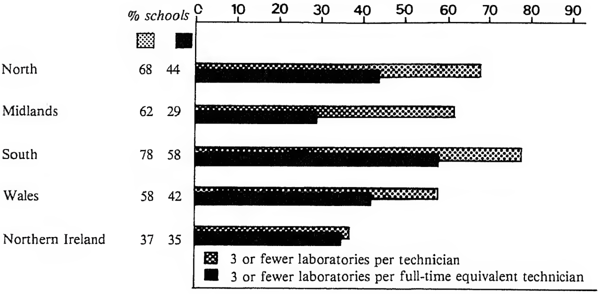
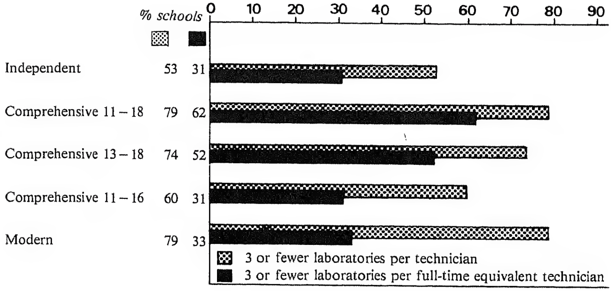


Figure 2.8 suggests that comprehensive schools with sixth forms have higher levels of technician support than all other types of school.

Figure 2.8 *Adequate level of technician support (% schools within type in England)*



The level of technician support was slightly higher in the metropolitan sample schools than in the non-metropolitan schools.

Overall, about 40% of the survey schools had adequate laboratory accommodation *and* adequate technician support for their laboratories. About 1 in 10 schools were *inadequately* provided in both respects.

In addition to laboratory work, it is recognised that schools can make use of a range of out-of-school facilities for science based activities. Schools were asked to indicate the types of out of school facilities used as part of the science programme for 15 year old pupils. About half of the survey schools had used field centres and/or had made industrial visits. Fewer than one fifth of the survey schools had used FE link courses.

2.2.3 Financial resources

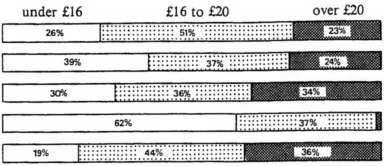
It is recognised that variation exists between local education authorities in the range and nature of items to be covered by the per capita expenditure allowances. Authorities may also change the number of items to be covered by the capitation allowance from one year to the next. For this reason, schools were not asked in the questionnaire for information about their capitation allowance. Instead, they were asked to give the amount spent by the school on teaching resources over a three year period. They were also asked for information on the school's science expenditure.

Expenditure on teaching resources

About one in five of the maintained sample schools either provided no expenditure information or could not provide it for all three years. Those schools which did provide complete data were found to divide roughly equally into the three ranges: *under £16, £16 – 20* and *over £20* per pupil per year.

The regional results are given in Figure 2.9. Within England, the survey schools in the South appear to have been receiving a generally higher level of financial support than those in the North and Midlands. However, the most notable feature of Figure 2.9 is the difference in the pattern of financial

Figure 2.9 *Average annual expenditure per pupil on all teaching resources (% schools within region)*

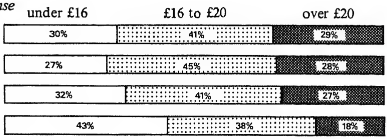


region	*no. of schools*	*% response*	under £16	£16 to £20	over £20
North	117	87	26%	51%	23%
Midlands	79	77	39%	37%	24%
South	141	74	30%	36%	34%
Wales	52	79	62%	37%	
Northern Ireland	36	71	19%	44%	36%

provision in the three countries, with proportionally more of the sample schools in Northern Ireland appearing in the high level group and proportionally more of the sample schools in Wales falling into the group receiving the lowest levels of financial support. If we look at the latest year alone (i.e. 1979/80) we find that the average per pupil expenditures for England, Wales and Northern Ireland respectively are £24, £16 and £26. Thus the general impression of Welsh schools being relatively under-financed in comparison to their English and Northern Irish counterparts still remains. It is perhaps pertinent to note that the apparent discrepancy in the level of financial support is not reflected in the official LEA statistical returns (see expenditure on books and equipment—CIPFA 1981). There is a widespread policy in Wales for certain capital equipment purchases to be made on behalf of schools directly by the LEAs, the schools concerned then receiving the *remainder* of their computed capitation allowances (WOED 1982); this could be a relevant factor.

Figure 2.10 provides the corresponding picture for the different types of maintained school in England, with apparently similar results for all three kinds of comprehensive school—the modern schools in general being less favourably financed.

Figure 2.10 *Average annual expenditure per pupil on all teaching resources (% schools within type in England)*



type of school	*no. of schools*	*% response*	under £16	£16 to £20	over £20
Comprehensive 11 – 18	133	90	30%	41%	29%
Comprehensive 13 – 18	64	94	27%	45%	28%
Comprehensive 11 – 16	75	82	32%	41%	27%
Modern	41	83	43%	38%	18%

There was little difference between the capitation findings for maintained schools in metropolitan and non-metropolitan locations.

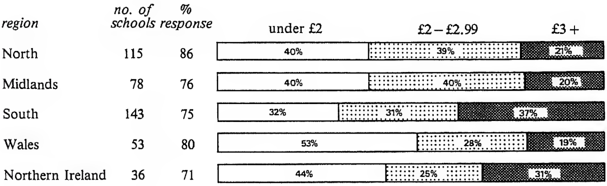
Science expenditure

Considerable variability exists between schools and local education authorities in the way funding is given to science departments, and what the allowance is to be used for. In some schools the department allowance has to cover not only the costs of apparatus and supplies for practical work but other running costs including those of stationery, reprographics, film hire charges, visits and field trips, whereas in other schools such non-specialist items are supplied from central funds. Such variability must be borne in mind when considering the levels of expenditure on science reported below.

As was the case for general expenditure on teaching resources, about one in five schools failed to provide details of their science departmental expenditure for all three years. For each school which did provide full information, an indication of its usual annual science expenditure was obtained by averaging over the three years the amounts spent from the general expenditure on teaching resources and any special grants received.

Schools were most often found to spend under £3 per pupil per year on equipment and other science teaching resources—a finding which agrees with that previously noted in the 1980 survey report for age 13 (DES 1982a).

Figure 2.11 *Average annual science expenditure: £ per pupil on roll (% schools within region)*



From Figure 2.11 we see that the survey schools in the South of England tended to spend more on science per pupil than did those in the North and Midlands. The survey schools in Wales appear in general to have been spending less on this area of the curriculum than those schools in England and Northern Ireland.

Figure 2.12 *Average annual science expenditure: £ per pupil on roll (% schools within type in England)*

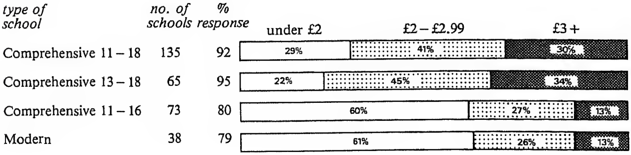


Figure 2.12 presents the results for the different school types in the maintained sector in England, and from this it seems that the 11 – 16 comprehensive most resembled the modern schools in terms of their levels of science expenditure. Despite the fact that the 11 – 16 comprehensives in general showed the same levels of expenditure on all teaching resources as the 11 – 18 and 13 – 18 schools, their science expenditure was lower (the difference presumably accounted for by sixth form costs).

There were no discernible differences in the science expenditure patterns shown by schools in metropolitan and non-metropolitan locations.

2.3

The pupils' science course choices

2.3.1 Information sought from the schools

Information about the science courses currently being undertaken by the sample pupils was gathered directly from the survey schools. Each school was asked to indicate which of the following science courses each selected pupil was taking:

Physics
Chemistry
Biology
Physics with chemistry
General science
Double certificate science or SCISP[1]
Human biology
Technology or engineering science
Electronics
Any other science course

The schools were also asked to indicate pupils who were not currently studying any science.

In addition to this curriculum information, the schools were further asked to classify each pupil into one of five groups according to the number and nature of the external examinations for which that pupil was likely to be entered in the following spring. The purpose of this information was to provide a classification of pupils into bands roughly indicative of varying levels of academic ability, so that differences in science subject choices across the ability range could be investigated.

Each pupil's sex was also recorded so that sex differences in science course choices would also be open to exploration.

The classification criteria and the proportions of the pupil sample falling into each group are given in Table 2.1.

Entries for the 16-plus common examination were omitted because without additional estimates of the grades pupils were likely to obtain, they give insufficient information to allow classification into these five groups.

It is recognised that this index is a crude one for the purpose, given that schools often have different and general entry policies for their pupils. It is

---

[1] SCISP refers to the course materials produced by the Schools Council Integrated Science Project.

Table 2.1 *Classification of pupil sample into five examination-entry groups*

Group	*Examinations to be entered*	*Proportion of pupil sample*		
		boys	*girls*	*all*
1	6 or more O-levels[1]	30	34	32
2	3 to 5 O-levels[1]	12	15	14
3	1 or 2 O-levels[1]	11	12	11
4	No O-levels but 5 or more CSE examinations	28	27	27
5	No O-levels and 4 or fewer CSE examinations	19	13	16

[1] with or without CSE's

also likely to be affected to an unknown degree as the common 16-plus examination becomes more widely adopted, thereafter being rendered irrelevant. In anticipation of this, mathematics and language performance results are being monitored for pupils in specific science curriculum groups by collaboration with the monitoring teams in these other areas. (Further details of this co-operative exercise will be found in the final chapter of this report).

According to the information contained in Table 2.1, 43% of these sample pupils were to be entered for CSE only or were not to be entered for any examination. This compares with a figure of 51% recorded for the 16 year olds surveyed by the National Child Development Study (Fogelman 1981). The discrepancy between these proportions could perhaps reflect the fact that more pupils are now being entered for examinations at both levels than was the case in 1974 when the NCDS sample pupils reached 16 years of age (this coinciding as it did with the raising of the school leaving age).

2.3.2 Overall patterns of science course choices

In contrast to the finding from the 1980 age 13 survey in science (DES 1982a), in which it was found that almost all of the survey schools offered their *13 year old* pupils a *common* science curriculum (separate subject courses in biology, chemistry and physics or, more usually, a single general science course), virtually all of the schools in this present survey sample appeared to offer their *15 year old* pupils a broad range of science subject options. Only 11 of the 568 survey schools offered a common science curriculum to all their 15 year old pupils—9 schools offering separate subject biology, chemistry and physics (five of these schools being independent or grammar schools), the remaining two offering general science.

Table 2.2 *The relative popularity of science course combinations*

Science course combination	*% pupils*
Biology only	19
Biology, chemistry and physics	11
Physics only	11
General science	11
Chemistry and physics	9
Biology and chemistry	6
Human biology	6
Biology and physics	5
Chemistry only	3
Other combinations	8
No science	11

About one in ten of the sample pupils were not currently studying any science at all at school.

As might be expected, of the 90% of pupils who were taking at least one science course, the vast majority (75%) were studying *either* general science *or* one or more of biology, chemistry and physics. Table 2.2 shows which subject combinations attracted at least 1% of the pupils, combinations being given in decreasing order of popularity.

If we look now at each science course separately and consider the numbers of pupils taking these in *any* subject combination, we see from Table 2.3 that roughly 40% of pupils were studying biology, a slightly lower proportion were studying physics and almost 30% chemistry.

Table 2.3 *The relative popularity of specific science subjects (subjects taken in any course combination)*

Subject	*% pupils*
Biology	41
Physics	37
Chemistry	29
General science	12
Human biology	9
Technology	2
Physics with chemistry	1
Double certificate science or SCISP	*
Electronics	*
Other science subject	8

*fewer than 1%

Most notable perhaps from Table 2.3 are the very low proportions of pupils who were taking courses in technology, electronics or integrated science.

2.3.3 Sex and ability differences

It is well-known that proportionally more boys than girls opt to study science once they are given a choice, and that proportionally more boys tend to take physics courses than do girls, girls tending rather to opt for biology or human biology (DES 1979, 1980a).

The same pattern of subject choice is evident in this survey sample, as is seen in Figure 2.13 which gives the findings for the main course combinations considered in the previous section.

It is relevant to note from Figure 2.13 that while 7% of the boys chose not to study any science at all, twice this proportion of girls were in the non-science group. In their recent secondary survey HMI found similar proportions of pupils who were not studying science (DES 1979).

Figure 2.13 *Sex differences in science course choices*

Science course combinations	*% pupils* boys	girls
Biology, chemistry, physics	14	9
Biology, chemistry	3	10
Biology, physics	7	4
Chemistry, physics	14	3
Biology only	7	33
Chemistry only	2	3
Physics only	18	3
General science	13	10
Human biology	2	11
Other combinations	12	–
No science	7	14

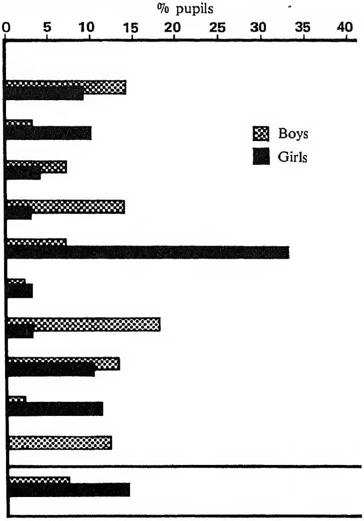


Figure 2.14 *Proportions of boys and girls studying each science course in any combination*

	% pupils boys	girls
Biology	31	52
Chemistry	34	25
Physics	56	19
General science	15	10
Human biology	4	11
Technology	3	*
Physics with chemistry	1	1
SCISP	1	*
Electronics	1	*

* fewer than 1%

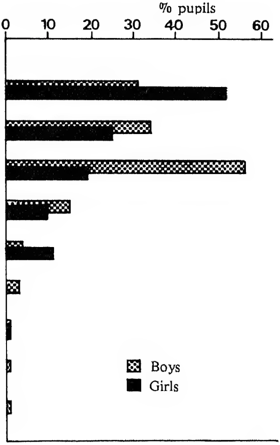
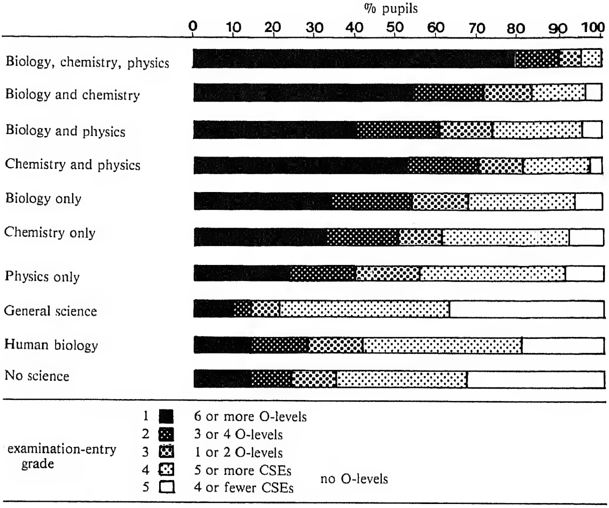


Figure 2.14 shows the proportions of boys and girls who were studying particular science subjects in any combination: the tendency for boys to study physics and girls biology is clear.

Figure 2.15 illustrates differences in the nature of the science courses taken by pupils in the different examination bands, with relatively large proportions of less able pupils taking a general science course rather than any combination of the three separate science courses of biology, chemistry and physics.

Figure 2.15 *Examination-entry grade distributions for different science course combinations*



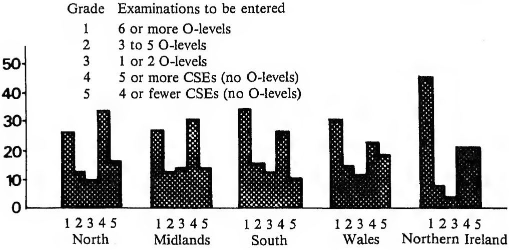
Despite the fact that the boys and girls were similarly apportioned into the five examination entry groups, the ability distributions are quite different for those groups opting for science course combinations involving physics. In particular, proportionally more of the girls than of the boys taking the major science options considered in this section were in the 'more able' group. For example, of those pupils taking physics (in any subject combination) about one-third of the boys were being entered for 6 or more O-levels, compared with *two-thirds* of the girls. For chemistry the difference is less marked with half of the boys in this high-entry group compared with three-fifths of the girls, while for biology the proportions were the same at two-fifths.

2.3.4 Regional and school type differences

As Figure 2.16 shows, the proportion of probable O-level candidates (ie. of sample pupils in examination entry groups 1 to 3) varied across regions, the North having the lowest proportion at 49%, the South the highest at 63%.

Within the probable O-level candidates, the distribution of pupils across the three groups was similar in the English regions and in Wales (an average of about 30% of all pupils in grade 1). The Northern Ireland pupil sample differed considerably from all other regional samples, with 46% of the Northern Ireland pupils being allocated to examination entry group 1 by their schools. It is probable that the Northern Ireland profile reflects the general entry policies of the schools in that region's totally selective system.

Figure 2.16 *Regional distributions across examination-entry grades*



The general *pattern* of science courses followed by pupils in the survey varied little from region to region. However, a major difference appeared between the regions in terms of the proportions of survey pupils who had *discontinued* a study of science in any form. The proportion of 'non-science' pupils was about 8% in all the English regions, rising to 11% in Wales. Among the survey pupils in Northern Ireland, however, a much higher proportion (26%) were no longer studying science. Consistently across all regions, the proportion of girls who were no longer studying any science was about twice as high as the equivalent proportion of boys. In Northern Ireland these proportions were 37% and 17% respectively.

As might be expected there were broad differences between the independent and grammar schools on the one hand and the comprehensive and modern schools on the other both in terms of their sample pupils' ability distributions and in terms of their patterns of science study.

Very high proportions of the sample pupils in the grammar and independent schools were being entered for 6 or more O-levels, the modern schools having the lowest proportion of such pupils.

The proportions of pupils not studying any science at all fell between 2% and 7% for the independent, the grammar and the comprehensive schools with sixth forms, the proportion rising to about 10% for the pupils in modern schools and 11 – 16 comprehensives.

While around 30% of the grammar and independent school pupils were studying all three traditional science subjects, under 10% of the comprehensive school pupils and only 2% of the modern school pupils were doing so. On the other hand, between 10% and 15% of the pupils in comprehensive and modern schools were taking a single general science course and similar proportions were studying physics only; under 5% of the independent and grammar school pupils were taking either of these courses. Fewer than one in a hundred of the comprehensive and modern school pupils were taking the SCISP course, this option not being found in the course choices of the sample pupils in independent and grammar schools.

2.4 Summary

Applying HMI's criteria (DES 1979) for judging the adequacy of a school's laboratory accommodation and technician support, it was found that about two-thirds of the survey schools had adequate numbers of laboratories to allow their pupils' science courses to be practically based, while a lower 40% of schools had adequate technician support for their laboratories.

Overall about one in ten schools were found to be inadequately provided both in terms of laboratory accommodation *and* technician support.

As far as financial expenditure on science education is concerned, schools were most often found to spend under £3 per pupil per year on equipment and other science teaching materials.

There were a number of resource differences for sample schools in the three countries. For example, proportionally more of the survey schools in Wales and Northern Ireland had inadequate laboratory accommodation and technician support than did those in England. Figures for expenditure on teaching resources revealed that the survey schools in Wales were less well-financed (at least directly) than those in England or Northern Ireland, and proportionally more of the Welsh schools spent the lowest amounts per pupil per year on science teaching resources.

In general, the survey schools in the South received higher levels of resource provision than did schools in other regions. For example, proportionally more of these schools had the lowest overall pupil/teacher ratios and the highest levels of expenditure on teaching resources. Further, proportionally more of them had adequate laboratory accommodation and technician support, and perhaps as a consequence of their general levels of expenditure were spending more per pupil per year on science education than were those schools in other regions.

There were few noticeable differences between the survey schools in metropolitan and non-metropolitan locations in England. The non-metropolitan schools were in general at a slight advantage in terms of adequacy of laboratory provision but the metropolitan schools in general had slightly higher levels of technician support. Proportionally more of the non-metropolitan schools had higher pupil/teacher ratios.

A very high proportion of the teachers who were teaching science in the survey schools held an appropriate qualification in this area, and there was a great variety of science course options available to 15 year old pupils in the schools;

only a very small minority of schools constrained their pupils to take a compulsory common science curriculum.

About one in ten of the survey pupils in England and Wales had chosen to discontinue a study of science in any form, the proportion of such pupils in the Northern Ireland sample rising to one in four. The majority of those still including a science component in their curriculum had chosen to study some combination of the three traditional science subjects biology, chemistry and physics or a single general science course.

The proportion of girls in the survey sample who had chosen to discontinue their study of science was twice as large as that of boys, and there is evidence of a tendency among those pupils who were still taking science courses for girls to opt for biology and boys for physics. Further, proportionally more of the girls than boys who were taking physics in any course combination were in the more able group.

3 Using symbolic representations

3.1 Introduction

Scientists frequently use symbolic representation to convey complex information in a precise and concise manner. Information presented in the form of graphs, charts and tables is common in much scientific writing, whether in textbooks or the popular press. As well as being able to read such information, pupils need to be familiar with the notations used in certain areas of science, particularly the physical sciences; circuit diagrams, magnetic fields and sectional drawings of glassware are typical examples.

Questions in this category have been divided into the following main groups:

sub-category 1α	Reading graphs, charts and tables
sub-category 1β	Constructing or completing graphs, charts and tables
sub-category 1γ	Using conventional representations.

The questions in sub-categories 1α and 1β contain enough information for answers to be given without recourse to any taught science concepts. The scores for the questions in each sub-category test are summed to produce a sub-category total.

Sub-category 1γ, on the other hand, is designed to assess pupils' knowledge of the notation of science, and hence is largely based on recall of taught information. The questions cover a disparate group of topics, and it was considered that the addition of scores from such varied questions would not give a meaningful total. A number of the questions are reported, either individually, or in the context of others on the same topic.

The aim of the next two sections is to acquaint the reader with the nature of the first two of the sub-categories, the scores from which are discussed in relation to the background variables in Chapter 9. Section 3.5 gives a more descriptive view of the questions in these sub-categories and, with this aim in mind, the questions have been grouped on a topic, rather than a sub-category, basis. This section also includes the results for a selection of the questions in sub-category 1γ.

3.2 Reading information from graphs, charts and tables: sub-category 1α

Thirty two questions from this sub-category were used in the survey. Six of the questions were concerned with reading data presented in a table, fifteen with graphical data and two with data in the form of a pie chart. Four questions involved schematic representation of processes or relationships and included items such as flow charts and Venn diagrams. A final group of questions required the pupils to label the axes of a graph.

Two questions from the sub-category 1α tests are shown, together with their numerical mark schemes, to illustrate how the marks are reduced to the scaled score for addition to the sub-category total. The contexts range from the sort of presentation familiar in science texts to questions based on material found in everyday settings such as newspapers and magazines. Detailed discussion of performance on individual questions is deferred to a later section.

In all the questions in this sub-category the demand being made is restricted to reading information from the tables, graphs and charts. The ability to make inferences or generalisations about presented data is considered elsewhere in the report.

Reading charts

Example question 1 (1α)

On the weather chart below these symbols are used:

✳ snow △ cloud ⊘ rain

▤ sunshine ◇ wind

Weather chart for 2 weeks.

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
▤ ◇	△	✳ △	✳ △	◇ ⊘	▤ △	▤ ◇
Day 8	**Day 9**	**Day 10**	**Day 11**	**Day 12**	**Day 13**	**Day 14**
✳ ◇	△	▤	△ ⊘	◇ △	▤ ◇	⊘

a) What was the weather on day 11?

. . . .

b) On which days was there sunshine but no wind?

. . . .

c) On which days was there cloud but no rain or snow?

. . . .

B1009 C 425

n = 772
Mean score 77%
Non-response 1%

Mark scheme

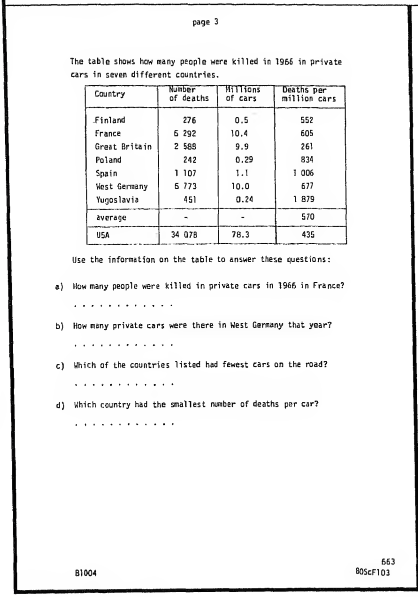
a.	Cloud and rain	1
b.	6, 10	1
c.	2, 6, 9, 12	1

Maximum score ③

Raw score	% pupils
0	2
1	19
2	24
3	55

Reading tables

Example question 2 (1α)



page 3

The table shows how many people were killed in 1966 in private cars in seven different countries.

Country	Number of deaths	Millions of cars	Deaths per million cars
Finland	276	0.5	552
France	6 292	10.4	605
Great Britain	2 588	9.9	261
Poland	242	0.29	834
Spain	1 107	1.1	1 006
West Germany	6 773	10.0	677
Yugoslavia	451	0.24	1 879
average	-	-	570
USA	34 078	78.3	435

Use the information on the table to answer these questions:

a) How many people were killed in private cars in 1966 in France?

. . . .

b) How many private cars were there in West Germany that year?

. . . .

c) Which of the countries listed had fewest cars on the road?

. . . .

d) Which country had the smallest number of deaths per car?

. . . .

B1004

663
80ScF103

n = 772
Mean score 67%
Non-response 4%

Mark scheme

a.	6292	1
b.	10,000,000 (10 million)	1
c.	Yugoslavia	1
d.	Great Britain	1

Maximum score ④

mapped score	raw score	% pupils
0	0	8
1	1	13
2	2, 3	50
3	4	29

3.3 Expressing information as graphs, charts and tables: sub-category 1β

As the questions in this sub-category require the pupils to complete or construct graphs, charts and tables, the number of questions which could be given in the time available was less than in sub-category 1α. Twenty six questions were used in the survey.

Seven of the questions required the pupils to complete a table, a schematic representation or a pie chart. A further eleven questions were concerned with the drawing of line graphs and bar charts. Seven of these required the pupils to construct the graph or bar chart from scratch, while four involved the completion of a graph on graph paper already labelled and provided with a scale. The remaining eight questions presented pupils with different graphical representations of a set of datum points and asked them to select the most appropriate.

Example questions 3 and 4 together with their numerical mark schemes illustrate the sub-category. Further examples of questions from both sub-categories will be found in section 3.5.

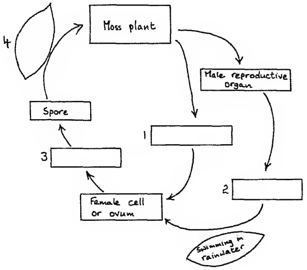
Completing a flow chart

Example question 3 (1β)

Read the description of the life cycle of a moss plant and then fill in the boxes labelled 1, 2, 3 and 4 in the diagram.

Mosses are small green plants which reproduce by spores produced in a capsule on a stalk. The capsule is formed from a female cell, or ovum, contained in a female reproductive organ. Before the ovum grows into a capsule it must be joined with a male cell, or sperm, which swims in rainwater from a male reproductive organ.

When the capsule is ripe it opens and the spores are blown about by air currents. Each spore which lands in a suitable place splits open and grows into a new moss plant. The life cycle then begins again.



C 24

n = 772
Mean score 35%
Non-response 13%

Mark scheme

1.	Female reproductive organ.	1
2.	Sperm (male cell).	1
3.	Capsule.	1
4.	Blown in air currents. (or similar)	1

Maximum score ④

mapped score	raw score	% pupils
0	0, 1	46
1	2	19
2	3	19
3	4	16

Completing a pie chart

Example question 4 (1β)

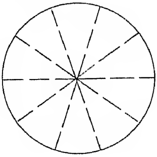
In the final years of the last century a survey was made to find out how many people were living in overcrowded homes in some areas of London.

The table shows some of the results of this survey:

No of persons per room	Percentage of total population living like this
3 or more	10
2 and under 3	20
1 and under 2	25
Less than 1	45

Use these figures to complete the pie chart below by shading or colouring.

Give a key to your chart.



B1064 C633

n = 772
Mean score 70%
Non-response 20%

Mark scheme

3 or more persons	– 1 segment
2 – 3 persons	– 2 segments
1 – 2 persons	– 2½ segments
0 – 1 persons	– 4½ segments

All correctly identified ①

3.4 Summary of scores: sub-categories 1α and 1β

The distribution of scores in the two tests was similar in shape but with more pupils scoring zero in the 1β test (3% compared with less than 1% for the 1α test) for which the mean was also lower. This lower mean in the 1β test is accounted for in part by the higher non-response rates found in questions requiring longer answers which occur mainly in the 1β test.

Boys obtained higher mean scores than girls in both tests but there was no obvious pattern to indicate which type of questions caused this difference.

The relationship between the test scores and certain background variables are reported and discussed in Chapter 9.

3.5 Description of performance levels on selected question types

A number of questions or groups of questions on particular topics or skills were marked on an extended category marking scheme and are reported in this section. A brief outline of the topics covered is given below.

Subsection 3.5.1	tables and pie charts.
Subsection 3.5.2	coordinate grids and line graphs.
Subsection 3.5.3	drawing line graphs from tabular data.
Subsection 3.5.4	schematic representation of electrical circuits and laboratory glassware.

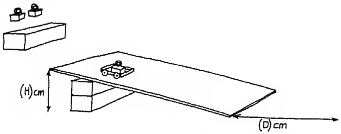
In the first three of these subsections, questions are drawn from sub-categories 1α and 1β. Subsection 3.5.4 deals with some questions drawn from sub-category 1γ.

3.5.1 Tables and pie charts

The simplest forms of such questions in which the pupil is asked to read data from or add data to a table have a high facility. Some 90% of pupils were able to deal with simple tables, the percentage dropping as the complexity rose. Example question 5 shown below proved to be the most difficult of this group.

Completing a table

Example question 5 (1β)



John and Pauline measured how far their truck travelled after it left the end of the board (distance 0 cm).
They changed the height of the end of the board (H cm).
They put different weights in the truck.

These are some of their results:

Height of raised end in cm (H)	Distance travelled in cm (0)				
	Nothing in truck	100 g in truck	200 g in truck	300 g in truck	400 g in truck
5	70	80	85		
10	85	95	103		
15	140	145	150		
20					
25					

a) They also got these results for the truck loaded with 300 g:

H = 5 cm   0 = 105 cm

H = 15 cm   0 = 160 cm

Add these results to the table above.

b) They put the height of the raised end up to 25 cm and found these results:

For 100 g in truck: 0 = 195 cm

For 200 g in truck: 0 = 200 cm

Add these results to the table above.

B 1093

C 431

n = 761

	% pupils	
	Age 13 n = 1042	Age 15 n = 761
All 4 correct	48	55
3 correct	5	3
2 correct	13	11
1 correct	3	6
0 correct	31	25
(no response)	(9)	(14)

Example question 6 is one of four involving pie charts, two asking pupils to read data from the chart and two asking them to complete one from data in the question.

Completing a pie chart

Example question 6 (1β)

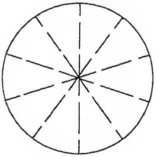
n = 772

The table gives the results of a recent survey of 100 people and shows how they chose to travel between London and Glasgow.

Method of travel:	Rail	Air	Bus	Car
Number of people:	30	40	20	10

(The figures have been rounded off to the nearest 10)

Illustrate these figures by shading or colouring the pie chart which has been drawn for you.
Give a key to your chart.



81065 C. 632

The maximum score was given when the four sectors were clearly defined by shading or colouring, and identified by label or key. Less complete responses received a proportion of the marks.

For pupils at age 13, 59% were able to complete the chart satisfactorily. At age 15, the proportion rose to 87%.

About 90% of pupils were able to complete the chart when sectors matched the ones given, as in the above question. If some interpolation was required, the success rate dropped to 70% with a rise in the non-response rate.

The questions on reading pie charts proved to be very easy. Over 90% of pupils were able to identify a particular sector of the chart as being the largest or smallest proportion of the whole.

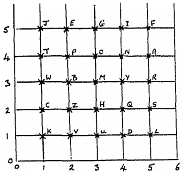
3.5.2 Coordinate grids and line graphs

A number of questions on reading, drawing or completing line graphs were available for analysis. Example question 7 gives some basic information on pupils' ability to read data located on a coordinate grid. As can be seen, between 70% and 80% of pupils were able to locate the coordinates of a point on a grid, but some 15% to 25% failed to use the normal convention.

Reading coordinates

Example question 7 (1*α*)

This is a way of making a code. The letters are replaced by the position of the cross near to them.
The code for M is (3,3).



What word does the following code make?

(3,2) (4,5) (1,4)

. . . .

C 357

n = 215

Response	% pupils
HIT	66
BAD	23
Other combinations showing inconsistent use of order; e.g. BIT, HAD, BID etc.	6
Answers involving incorrect letters or incomplete answers	3
No attempt	3

Discussion

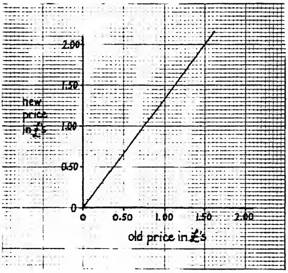
Two questions involved location of points on a simple grid. In addition to the one shown above, a further question required the pupils to locate points and join them up to produce a particular pattern on the grid. The data from both questions is summarised below.

	Question 7 % pupils	Further question (not shown) % pupils
conventional use of coordinates	66	55
reversal of convention	23	15
inconsistent use of convention	6	6
incomplete answers	3	21
no attempt	3	3

In Example question 8, similar performance levels were found for pupils reading from a graph. When the point to be read involved finding the x coordinate, and reading the y-value when both points are on a printed line of the graph paper, more than 90% of pupils were successful.

Reading line graphs

Example question 8 (1α)



This graph shows new prices compared with old prices.

a) What is the new price of something with an old price of £1.50?

. . . .

b) What was the old price of something with a new price of £1.20?

. . . .

c) What is the old price of something with a new price of £1.00?

. . . .

C 356

n = 223

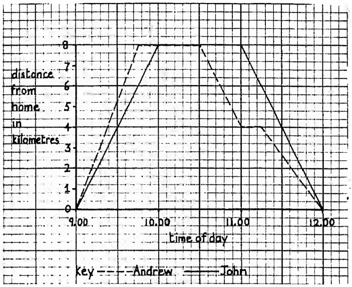
	% pupils
a. £2	91
b. £0.90	57
c. £1.75	72
all correct	49
no attempt	3

Reading a point occurring on the grid lines of the graph can be successfully accomplished by 91% of pupils, while points requiring interpolation on one or two axes become progressively more difficult.

Another factor affecting the scores is the degree of interpretation involved in the question. In Example question 9 the burden of the question was not simply one of reading datum points. Some minimal level of interpretation or knowledge is involved which was reflected in the percentage success on the parts of the question. A more detailed discussion of questions involving interpretation of data is given in Chapter 6.

Reading line graphs

Example question 9 (1*a*)



Andrew and John set out from home on their bikes at 9.00 to go to the swimming baths 8 kilometres away.
This is a graph of their journeys.

a) How long did John take to get to the baths? . . . .

b) Who stopped on the way home? For how long?

. . . . stopped for . . . .

C 423

n = 223

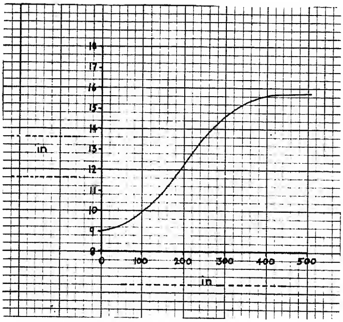
	% pupils
a. 1 hour	61
b. Andrew	83
c. 15 minutes	52
all correct	38
no attempt	7

Four of the questions on line graphs asked pupils to label the axes of the graph (see Example question 10). This type of question caused some problems for most pupils. The non-response rate was high, perhaps as a consequence of the relative unfamiliarity of this type of question. Of those who started the questions, many picked up clues from the question stem and reproduced these as their answers. Often this meant that instead of the two labels required, the quantity and the units, only one was quoted, usually the one which appears to be most familiar. For example, in the question shown, the unit of length is quoted far more frequently than the quantity itself. It must be assumed that some pupils are unaware of the necessity to quote both, assuming that 'cm' is just as good as 'length in cm'.

Labelling axes

Example question 10 (1*a*)

The graph shows how a spring stretches when bigger and bigger loads are hung from it.



Label the axes of the graph by filling in the correct words on the dotted lines.

C 639

n = 223

	% pupils
Length	28
cm, mm, m, or other suitable unit.	43
Mass or weight or load	51
g, kg, or imperial units.	41
Length AND mass, weight or load	27
cm, etc. AND g, kg etc.	38
Length in mm, etc. AND mass etc. in g, kg.	23
Axes reversed	7
No attempt	34

3.5.3 Drawing line graphs from tabular data

Seven of the questions on line graphs presented data in the form of a table and asked the pupils either to complete a graph on axes already drawn and labelled, or to construct a complete graph. The marking of the first of these types of question concentrated on skill in plotting, while that for the second type was mainly concerned with the pupil's ability to set up and label suitable axes.

Plotting

An example of the first type of question is given in Example question 11. A second question gave a table of data concerning crop damage over a number of years. The third and last of this type gave a table of temperatures and times for the cooling of two liquids.

The points were identified as occurring on a line of the graph paper or requiring interpolation between such lines. The accuracy of plotting was then assessed by coding each point as being within certain ranges as shown on the mark scheme of the example question.

About three quarters of the pupils were able to plot two thirds of the points occurring on a line, this fraction falling to a half when the points involved some interpolation on one or both axes. (The graph paper was 2mm pitch.)

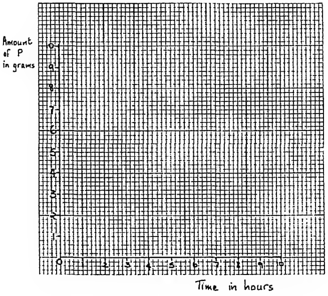
Completing a line graph

Example question 11 (1β)

The table shows how a substance, P, gradually changes to another substance during a chemical reaction:

Time in hours	0	2	4	6	8	10
Amount of P left in grams	8.0	6.4	5.3	4.4	3.7	3.2

Plot a graph to illustrate these figures.



C635

n = 717

Component of response	% pupils
a.	
3+ points on a line (i.e. requiring no interpolation	62
Points requiring interpolation	55
b.	
Either all points within ± ½ division	49
or 4+ points within ± ½ division	21
or 4+ points within ± 1 division	5
or 4+ points within ± 3 divisions	9
c.	
Either smooth curve (not necessarily going through all points	33
or smooth curve going through all points	18
or straight line segments	23
Bar chart	4
No attempt	8

The overlay used to mark the points allowed for an accuracy of ¼ division. Data was collected for lower accuracies of ½, 1 and 3 divisions for all the points and for 2/3 of them.

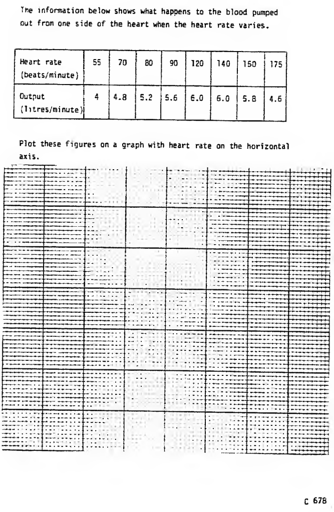
All three questions gave results which followed the pattern given above.

Four questions presented pupils with a sheet of graph paper and asked them to construct a graph from a table of datum points which were chosen to fit sensibly on the sheet. Each axis was marked separately on identical mark schemes.

The data from one such axis is shown for Example question 12, and is typical of the responses for three of the four questions. (The fourth will be discussed separately.) In general, pupils who successfully set up an equal interval scale (about 65%, range 57% – 74%) were also successful at selecting a sensible scale. About half (range 45% – 53%) labelled the axes adequately, 40% going on to put units on the axes (range 22% – 50%). Of the third who failed to set up an interval scale, half of them made no attempt while the rest made a variety of errors, the most common one being to plot the values from the table on adjacent grid lines of the graph.

Constructing a line graph

Example question 12 (1β)



The information below shows what happens to the blood pumped out from one side of the heart when the heart rate varies.

Heart rate (beats/minute)	55	70	80	90	120	140	150	175
Output (litres/minute)	4	4.8	5.2	5.6	6.0	6.0	5.8	4.6

Plot these figures on a graph with heart rate on the horizontal axis.

C 678

n = 717

	Component of responses	% pupils
a. Y axis	equal interval scale	71
	sensible choice of scale	64
	correctly labelled (output)	52
	correct units (litres/min)	41
b. X axis	equal interval scale	59
	sensible choice of scale	56
	correctly labelled (heart rate)	50
	correct units (beats/min)	39
c.	6 or more ± ½ division	46
	Either smooth curve (i.e. not necessarily going through all points	13
	or smooth curve going through all points	20
	or straight line segments	23
	Bar chart (if axes reversed as well, code as B)	4
	Axes reversed	13
	No attempt	22

Four questions were of this type, requiring pupils to draw a line graph on their own choice of axes. All the questions were marked using the scheme shown above. A scale was considered to be 'sensible' if it occupied more than half of the graph paper. In part (c) credit was given if approx. 2/3 of the points were correctly plotted.

The most common error in labelling the axes was to omit the quantity or the units on one or other of the axes. Even in Example question 12 in which the complete labels were given in the data table, a significant minority of pupils used the quantity only, failing to include the units. The same pattern of errors was encountered in Example question 10 discussed earlier. About 5% of pupils drew a bar chart rather than a line graph, even in questions which contained the explicit instruction to 'draw a line graph.'

Drawing

Information on how pupils connected up the points was available for all the questions. The data showed a fairly consistent pattern and is summarised in Table 3.1. About a third of the pupils either failed to start the question or connected the points up in a non-conventional fashion, a third used straight line segments and the remaining third drew a curve through all the points.

In one of the questions, one point was chosen to be clearly away from the line showing the general trend in the data. As can be seen from Table 3.2, very few pupils drew a line to illustrate this trend (8%), the majority drawing a line through all the points.

Table 3.1 *Drawing: data averaged for six questions*

Nature of response	average % pupils	range (%)
No response	17	7–27
Drawing not possible to fit to a category	27	16–37
Straight line segments	29	20–35
Smooth curve	38	28–45

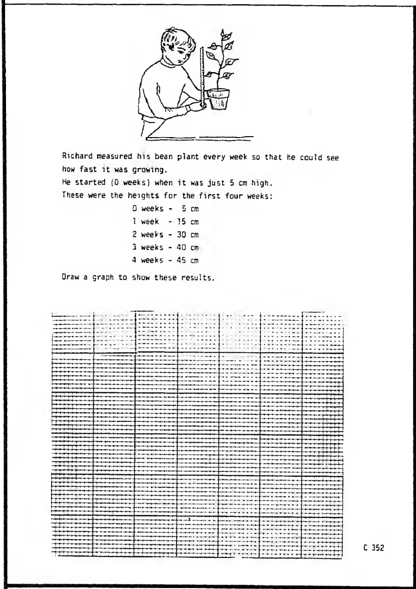
Table 3.2 *Data for question where one point did not lie on the smooth curve.*

Nature of response	% pupils
No response	15
Drawing not possible to fit to a category	9
Straight line segments	35
Smooth curve showing the trend in the data	8
Smooth curve through all points	33

Constructing a line graph

Example question 13 (1β)

n = 650



Richard measured his bean plant every week so that he could see how fast it was growing.
He started (0 weeks) when it was just 5 cm high.
These were the heights for the first four weeks:

0 weeks - 5 cm
1 week - 15 cm
2 weeks - 30 cm
3 weeks - 40 cm
4 weeks - 45 cm

Draw a graph to show these results.

C 352

The non-response rate for this question was much lower than for the other questions, 7% compared with 17%, a fact which may throw some light on the differences in performance. In contrast to the other questions where few pupils drew a bar chart, 23% did so for this data. The pattern of responses on setting up the axes was also different, over 80% of pupils setting up sensible interval scales, 70% putting units on the axes, but only 13% labelling the axes with the quantity being measured. (i.e. length and time.)

The figures suggest that the 10 per cent more pupils who attempted this question were those who were familiar with this type of graph, perhaps from their earlier years at secondary school, where it may have been presented as a bar chart for simplicity and where the stricter rules regarding the labelling of axes may not have been encountered.

Discussion

Because of the differing requirements of the mark schemes at the different ages, it is not possible to make direct comparisons. However, the results at age 15 can be interpreted in the light of the category marking scheme used at the other ages. Such an interpretation suggests that about 50% of pupils at age 15 meet the criteria, compared to 24% at 13 and 15% at 11.

Criteria for marking of question 13 (ages 11 and 13)

1. The direction of the axes is conventional.
2. A linear scale is used on both axes.
3. A sensible choice of scale is made on both axes.
4. The axes are sufficiently well labelled to be identified by quantity or unit or both.
5. At least three points correctly plotted.
6. The points are connected by a line, implying that one quantity varies continuously with the other.

As mentioned earlier, one of the four questions, Example question 13, did not follow the pattern of the remaining three. This question is shown with its mark scheme and a comparison with equivalent data from the 13 year old survey in which it was also used.

Summary of question results: sub-categories 1α and 1β

The demands of most of the questions in sub-categories 1α and 1β relate to a small number of basic skills. Pie charts involve matching a fraction to the size of a sector; tables, grids and line graphs rely on the skill of coordinate location, with other questions being either extensions of these skills or complementary to them.

The most basic skills appear to be within the grasp of most pupils, some 90% being able to read a pie chart and a coordinate grid or line graph where no interpolation is required. Any qualification of that skill results, not unnaturally, in increasing the failure rate—interpolation on two axes of a graph, for instance, increasing it to 50%.

In the longer questions on constructing graphs, about half of the pupils tested were able to set up axes, plot most of the points in approximately the right places, and join them up with straight lines or curves. It is at the more detailed level that failure occurred in such things as accuracy of plotting, correctness of labelling the axes and techniques of drawing.

3.5.4 Schematic representation of electrical circuits and laboratory glassware

This section presents the results of four questions from the 1γ sub-category. The first two were concerned with the use of conventional diagrammatic representation of circuits. Example question 14 is typical of the sort of problems encountered early in secondary school.

The results show that more than half the pupils tested were familiar with the symbols for a battery, switch and bulb; rather less than half being familiar with that for an ammeter. However, when asked to convert the pictorial representation into a circuit diagram, the percentage success rate on each component dropped by about 10%.

About a third of the pupils were able to draw a satisfactory circuit diagram, but only 17% used the conventional symbols for all the components. There was a large non-response rate on this question (34%).

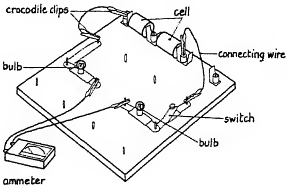
The final two questions dealt with the drawing and labelling of glassware. In Example question 15, only 7% of pupils were able to name all the items of apparatus accurately although about 40% were able to give a sensible general name to all the components.

In Example question 16, pupils were asked to turn a pictorial 3D representation of a simple piece of apparatus into a section drawing. The results are presented alongside the question and show that about a fifth of the pupils were able to make a reasonable section drawing.

Schematic representation of electrical circuits

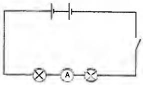
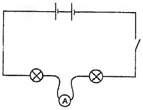
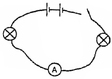
Example question 14 (1γ)

Here is a labelled drawing of a circuit.



Draw a circuit diagram instead, using the proper symbols:

C 673

		n = 223 % pupils
A.	Mark if the following are shown:	
	1 cell	1
	2 cells	41
	1 bulb	1
	2 bulbs	48
	Ammeter	46
	Switch	45
	Diagram matching drawing	33
B.	Mark if accepted symbols are used for	
	cells —|⊢—	39
	bulbs or ⊗	32
	ammeter Ⓐ	34
	switch —/—	40
	All correct symbols	17
C.	Categorise general shape of diagram according to the diagrams below	
	A 	40
	B 	13
	C 	2
D	Diagrams with some components in parallel	2
E	Other forms of diagram	10
	No attempt	34

Discussion
Another related question required the pupils to give names to symbols in a conventionally drawn diagram of a series circuit. The results of this question are shown below.

Symbol	Mark scheme	% correct on each point
—|⊢—	Battery or cell	57
—/—	Switch	51
	Bulb	59
Ⓐ	Ammeter	45
No attempt		22
All correct		30

n = 208

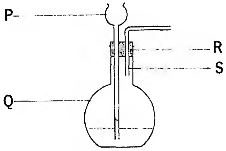
The mean scores for both these questions showed boys scoring almost double that for the girls.

Schematic representation of laboratory glassware

Example question 15 (1γ)

The apparatus in the diagram below has been labelled P, Q, R and S.

Write down - against the letters underneath the diagram - the name you would give each piece of apparatus.



P . . . .

Q . . . .

R . . . .

S . . . .

B 1087

C 620

n = 203

Mark in 2 categories.

A. use of correct terms.

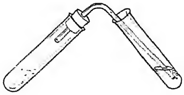
B. use of more general but acceptable terms.

			% pupils
A.		(Correct terms)	
	P	THISTLE FUNNEL	36
	Q	FLAT BOTTOMED FLASK	17
	R	BUNG	43
	S	DELIVERY TUBE	15
		ALL CORRECT	7
B.		(General but acceptable terms)	
	P	FUNNEL (etc.)	7
	Q	FLASK (or other term including the word flask)	37
	R	CORK, STOPPER etc.	41
	S	TUBE, GLASS TUBE, etc.	51
Scored zero (no attempt)			15(10)

Schematic representation of laboratory glassware

Example question 16 (1γ)

Look carefully at the drawing below.



Make a section drawing of the apparatus, set up as shown in the diagram above.
Use the space below.

B1077

C 630

n = 203

Approximately three quarters of the pupils attempted this question, a fifth producing a true section drawing. Of the remainder, the errors were mainly ones of detail: bungs drawn in three dimensions, absence of liquid levels in the boiling tubes and presence of a liquid level in the delivery tube being most common. 10% drew a diagram of the dismantled components.

3.5.5 Summary of question results: sub-category 1γ

The mean scores on the 1γ questions were low and accompanied by high non-response rates. In contrast to the sub-category 1α and 1β questions where means and non-response rates were, on the whole, greater than 50% and less than 15% respectively, the figures for the 1γ questions showed means below 50% and non-response rates typically 20% – 30%.

It is notable that, even though the majority of pupils are introduced to electricity in their first two years of secondary education, only about half of them recognise the symbol for a battery and less than a fifth can draw a circuit diagram correctly. These results must be considered in the light of the fact that almost 50% of pupils study physics or a related subject in their fifth year of secondary education.

In a similar vein, about a third of pupils study chemistry, yet only a fifth or less can draw and label diagrams of laboratory glassware accurately.

The contrast between the scores on this group of questions and those in the 1α and 1β tests is marked. Skills of data handling are better developed than the more content laden and specific techniques tested in the 1γ questions.

4 Using apparatus and measuring instruments

4.1 Introduction

The effective use of measuring instruments and apparatus is commonly seen as an important outcome of science teaching. The skills that are required include a basic manipulative ability, a careful and orderly method of observing and recording events, and an understanding of measurement concepts such as quantity, interval scales and units. The ability and willingness to follow instructions with an appropriate degree of initiative and caution is also required.

In order to develop practical skills, pupils are often trained in the use of specific instruments such as thermometers, lever arm balances and rulers. One method of assessing the levels of performance of pupils in practical skills would be to measure the degree of skill shown on each of these specific instruments. Over 150 science teachers from different disciplines, schools and regions of the country assisted in the compilation of a list of instruments and apparatus that could be used in this assessment. (This list is given in Appendix 1.) They also suggested laboratory procedures that 15 year old pupils could be expected to know by the time they were ready to leave school. (These are listed in the same Appendix.) Information on the performance levels by the pupils on some of the instruments is given in the description of some of the questions.

An alternative approach to assessing practical skills is to identify aspects of instrument use that are of general importance and then test them using a variety of instruments. For example, being able to give the value that lies between two of the calibration marks indicated on the scale of an instrument might be regarded as important. This skill could be measured using a ruler, a measuring cylinder or a thermometer. Whilst being able to read the volume of liquid in a 100 cm^3 measuring cylinder might be of little value outside a school laboratory, being able to make a reading on a scale using a liquid meniscus as a marker could well be in such different contexts as following a recipe in the kitchen and on the factory floor using a hydraulic pump. In view of the rapidly changing nature of scientific instrumentation it is at least as important for pupils to be familiar with the general characteristics of a range of measuring instruments as with the idiosyncracies of a few.

Three distinct aspects of practical skills were identified:

sub-category 2α	Using laboratory instruments — pupils are required to use units, scales, measuring and laboratory instruments correctly.
sub-category 2β	Estimating — pupils are required to estimate the magnitude of physical quantities by judging their value, or by attempting to take a specified amount of material.

sub-category 2γ Following instructions
— pupils are required to follow written instructions and/or diagrams. Following the instructions may require simple observation, recall of procedures or exact compliance with a set of directions.

In what follows, each sub-category is discussed in more detail and illustrated with example questions. Comments about the degree of competence with which pupils used the specific instruments are made in the discussion of the questions, while pupils' performances on the more general skills are summarised at the end of each sub-category. The first section briefly outlines how the tests were administered.

4.2 Administration

The reliability of the practical tests would not be adequate for a national survey of this kind if the variation which might arise from the use of a school's own apparatus and staff resources was not eliminated. To ensure greater uniformity in the administration of the tests trained administrators travelled to each school, taking with them standard sets of all the apparatus required.

All schools which participated in the practical testing agreed to release a laboratory for one day and to provide a science teacher who observed the pupils as they attempted one of the questions.

Each test was set up as a "circus". That is, the questions were laid out around the laboratory at positions called "stations". There was one pupil at each "station"; after six minutes the pupils were all moved on one station and so each pupil saw every question but only one set of apparatus per question was used. Usually there was only one question but occasionally there were up to three short questions at a station. It was necessary for each pupil to be presented with the apparatus in the same state as the pupil before. This meant that some stations needed to be "serviced", ie. the apparatus needed to be reset to the original layout.

In total 155 schools were visited and 2,790 pupils tested. Each question was attempted by 558 pupils drawn from 62 different schools. 29 questions in this category were used in this survey.

To ensure that all pupils had a reasonable understanding of how the test would be carried out, the administrator explained how the circus had been set up and how pupils were to move from station to station. The printed rubric shown below was given to pupils as an aide-memoire and no problems were encountered in the administration of the test. The administrators reported that many of the pupils seemed unfamiliar with a circus organisation for practicals.

The rubric

These questions are designed to find out how good you are at doing practical work and observing.

* You will have to move from station to station to see all the questions.
* Your card tells you which station to go to when the supervisor tells you to move.
* At each station you will find a pile of questions.
  Take one and answer it.
  Join this question to this piece of paper with the tag.
* You may ask for the supervisor if you need help.

More detailed information about the administration of the practical tests can be found in Appendix 3.

4.3 Using laboratory instruments: sub-category 2α

4.3.1 Outline and description of questions

In what follows the intent of the sub-category is explained principally by describing the characteristics of the questions that were given to pupils. Four questions are then discussed in some detail and finally the overall results for this sub-category are given.

In this sub-category all the questions require an instrument to be used correctly, ie. according to an established procedure. There is a correct way of using a ruler to measure length, and of using an ammeter to measure current. If this procedure is not adopted, error in the measurement is likely to occur.

The questions range from multiple-choice questions on knowledge of units, through questions on reading a preset value on a scale, to the use of an adequate measuring procedure to determine small values of distance or time.

Questions such as Example question 1, which test knowledge of units were given as the first of two questions at a station. The second question would be one that took no more than 5 minutes to complete. In the question shown, 64% chose the correct unit for the forcemeter scale. Of the six questions of this type that were used in this year's survey the range of mean scores was 55–82%.

Units of force

Example question 1 (2α)

n= 557
Mean score 64%
Non-response 4%

Which units would you find on a forcemeter?

Tick in the box next to the answer you think is correct.

☐ A cm^3
☐ B g
☐ C N
☐ D S
☐ E cm

Response	% pupils
A	11
B	8
* C	64
D	8
E	3
Multiple response	2
Non-response	4

Maximum score ①

In general those questions that asked pupils to choose the unit to go with the instrument had higher mean scores than those that asked for the instrument that went with the unit.

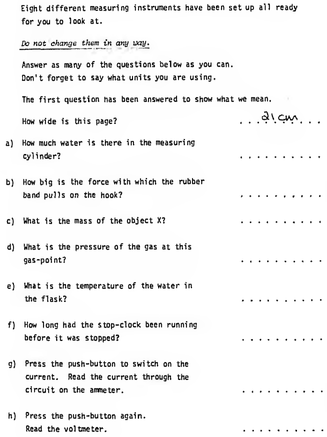
Another type of question tested the pupil's ability to read scales on instruments. Several different instruments were used and they were pre-set at a particular value. In this way, failure by a pupil to score could not be due to using the instrument incorrectly. Marks were only awarded if the pupils gave a value and the unit. If the value lay within a predetermined range that was appropriate to the accuracy of the instrument then the pupil scored 2 marks. If the value given lay within a range about double that of the pre-defined range the pupil scored 1 mark.

Example question 2 illustrates this type of question.

A similar question is reported in the age 13 report (DES 1982a). The results are not *directly* comparable, however, because the mark schemes differ. The criteria used to determine whether a performance is adequate at 13 may differ from that used at age 15. This is reflected in the ranges of values allowed at the two ages for the various instruments.

Reading scales

Example question 2 (2α)



Eight different measuring instruments have been set up all ready for you to look at.

Do not change them in any way.

Answer as many of the questions below as you can.
Don't forget to say what units you are using.

The first question has been answered to show what we mean.

How wide is this page? . . .21cm. . .

a) How much water is there in the measuring cylinder? . . . .

b) How big is the force with which the rubber band pulls on the hook? . . . .

c) What is the mass of the object X? . . . .

d) What is the pressure of the gas at this gas-point? . . . .

e) What is the temperature of the water in the flask? . . . .

f) How long had the stop-clock been running before it was stopped? . . . .

g) Press the push-button to switch on the current. Read the current through the circuit on the ammeter. . . . .

h) Press the push-button again. Read the voltmeter. . . . .

n = 558
Mean score 57%
Non-response 0%

Mark scheme

	Value	Units	Narrow Range 2 marks	Wide Range 1 mark
a.	78	cm^3	± 0.5	± 1
b.	40	N	± 5	± 10
c.	116	g	± 1.5	± 2.5
d.	13	cm	± 3	± 5
e.	44	°C	± 1.5	± 2.5
f.	7:17	min:s	± 0	± 1
g.	0.27	A	± 0.01	± 0.02
h.	3.4	V	± 0.1	± 0.2

Code U for any part where units omitted

Maximum score ⑯

mapped score	raw score	% pupils
0	0, 1, 2	4
1	3,4, 5, 6, 7	31
2	8, 9, 10, 11, 12	56
3	13, 14, 15, 16	10

Each of the parts of this question was marked and coded up separately so that the performance on each type of instrument could be analysed. The results are shown in Table 4.1 (p. 48). The instruments in the table have been ranked in order of difficulty.

Table 4.1 *Reading scales on laboratory instruments*

		Calibration		% of pupils		
Instrument	Set at	numbered values on scale	sub-division on scale	within narrow range	within wide range	failing to score
Thermometer	44°C	10°C	×1°C	82	4	14
Forcemeter	30 N	10 N	×2 N	82	2	16
Voltmeter	3.4 V	1 V	×0.1 V	75	5	20
Stop clock	7 min 17 s	5 s	×1 s	59	2	39
Lever arm balance	118 g	10 g	×1 g	38	14	48
Measuring cylinder	78 cm^2	10 cm^3	×2 cm^3	22	28	50
Manometer	13 cm	1 cm	×1 cm	19	6	75
Ammeter	0.27 A	0.1 A	×0.02 A	15	2	83

Just under 4% of the pupils lost marks because they omitted the units. No pupil consistently read the instrument correctly *and* failed to give units. There were, however, a few (about 5%) who gave no units at all. These pupils usually gave values that were outside even the wider range allowed.

Apart from the manometer, the difficulty experienced by the pupils in giving a correct reading appeared to be influenced more by the calibration of the scale and whether the value indicated was on a major, numbered division, than to the type of instrument — or its familiarity. The problems encountered with the manometer were probably due to the need to find a difference rather than merely read off a value, which is something the pupils would have to know.

The level of performance on the questions that tested the ability to read a value off a scale was found to be consistent over all four questions of this type that were used this year.

The last group of questions that were included in this sub-category are ones that required the pupils to use instruments in standardised procedures. To know how to use an instrument correctly involves knowing how to employ the appropriate procedure.

13 questions of this type were used this year, two of them are given as examples.

In Example question 3 the pupils had to use a lever arm balance to make four weighings. There was only a single scale on the balance which ran from 0–25 g (marked off in 1 g sub-divisions). Every tenth gram was numbered.

Weighing

Example question 3 (2*a*)

n = 558
Mean score 31%
Non-response 2%

This question is to find out how good you are at weighing.

If the towers stick ask the supervisor for help

- Weigh out exactly 50 g of sand.
- Put it into the top of the tower of sieves.
- Put the lid on the tower.
- Hold the tower upright and tap its side to shake the sand down. Count slowly to 20.
- Carefully remove sieve number 1.
- Pour the sand into the weighing pan.

a) The sand left in sieve 1 weighed . . . . g

- Tip this sand into the waste box.
- Put the lid back onto the tower.
- Carefully remove sieve number 2 and weigh the sand.

b) The sand left in sieve 2 weighed . . . . g

- Tip this sand into the waste box.

c) The sand left in sieve 3 weighed . . . . g

Mark scheme

Contents of sieve:	
Sieve 1: 4–6 g	1
Sieve 2: 16–24 g	1
Sieve 3: 24–26 g	1
If total of weights lie between 48.5–50.5 g	1
Maximum score	④

mapped score	raw score	% pupils
0	0	41
1	1	35
2	2	16
3	3, 4	8

The difficulty experienced by pupils when they were asked simply to read a value off the scale was probably responsible for the generally low performance found on this question. If the pupils failed to measure out 50 g accurately in the initial stages of the task they would then find it difficult to score any marks for the rest of the question. In many cases when the values that were given as answers to the three parts by pupils were summed they gave total weights very different from 50 g. (They ranged from 20–120 g.)

About 8% of the pupils failed to finish the task completely. In the main these scored zero because the one or two values they did give were outside the range of acceptable values, suggesting that they were having difficulty with the task.

Some of the aspects of the weighing procedure that were not carried out correctly included: failing to centre the pan on the arms, failing to empty the pan completely before the next weighing and failing to prevent the lever arm balance from sticking. (When using a lever arm balance of this type part of the procedure is to press the pan down a little and then let go so that the pointer moves beyond the value at which it will eventually stop.)

Example question 4 required the pupils to use a technique for measuring a small time interval: the period of a simple pendulum. The appropriate technique, using the apparatus provided, would be to time several swings and calculate a mean value.

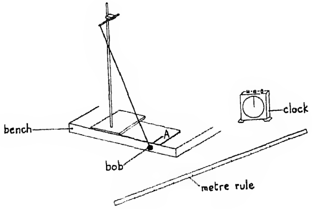
Great care was taken in wording the question so that it would not suggest a technique.

This was found to be one of the hardest of the repeated measurement procedure questions. 57% of the pupils thought it reasonable to attempt to time one swing once. 36% did, however, use a method based on timing a number of swings. Four questions of this type were used in the survey. Some asked for very small distances to be measured or an average value to be found (mean scores ranged from about 10% to 70%).

Repeated measurement

Example question 4 (2*a*)

n = 557
Mean score 24%
Non-response 1%



- Think of a way of making the following measurements as accurate as possible
- Find the time it takes for the bob to swing from A and back again.
- Use the space below for any working out you want to do.
- Say how you took the measurements . . . .
- The length of time for swinging is . . . .

Mark scheme

Measures 1 swing only	0
Repeated measurements of single swing	1
Counted swings in a fixed period of time	2
Timed fixed number of swings	3
Swing time: 1.15 (second)± 0.15 (S)	2
1.15 (second)± 0.25 (S)	1
*No score if the only answer on the script is 1 sec.	
If 1.0 sec mark normally (Units are not essential)	
If mathematics correct but arithmetic wrong	1

Maximum score ⑤

mapped score	raw score	% pupils
0	0	64
1	1	9
2	2, 3	15
3	4, 5	12

4.3.2 Summary of 2α performance

24 questions in the sub-category were used in this survey, eight in each of three circuses.

The mean performance level for this sub-category as a whole was around 40%, with a statistically significant difference between the performance levels of boys and girls (43% and 38% respectively).

About 1% of pupils failed to score on this sub-category. In general, there was a low non-response rate and most pupils appeared to enjoy the tests, feeling that they were able to cope adequately with the tasks.

A major problem that was uncovered was the difficulty many pupils had when asked to read values off instrument scales. This was particularly evident when the value was one which was not specifically marked on the scale. This suggests that either more care should be taken in the way in which instruments are calibrated for use in school laboratories, or more attention should be paid to teaching pupils how to read a variety of scales.

Certain types of failure concerned with using appropriate procedures with the instruments also caused the pupils to lose marks.

It was found, however, that pupils were likely to include the units as well as the value in recording a measurement if they could read the instrument accurately.

4.4

Estimating: sub-category 2β

4.4.1 Outline and description of questions

Before using any instrument it is important to know whether its range is suitable. This requires a guess as to what is the order of magnitude for the value of the quantity to be measured. All the questions ask for considered "guesses" about the familiar quantities of length, area, volume, mass, force, temperature and magnification in the ranges that are typically encountered in laboratory situations. Having a "feel" for the magnitude of quantities is valuable when dealing with both practical and theoretical aspects of science. When reading books or work sheets it is expected that references to the magnitude of a quantity will be meaningful. Scores in this sub-category throw light on just what meaning is conveyed.

The questions in sub-category 2β either required the pupils to give values to quantities supplied to them or, using given materials, to set up and leave for marking quantities they considered close to the values given.

There was only one station in each circus but several quantities had to be estimated at the station. Example question 5 illustrates this general type.

Estimating

Example question 5 (2β)

n = 557
Non-response 0%

You are to make your best guess at the answers to these questions.
You are not allowed to use measuring instruments.

a) What is the length of the rod labelled A? . . . . cm

b) What is the mass (or weight) of the rod labelled A? . . . . g

c) What is the volume of the bottle labelled B? . . . . cm^3

d) What is the area of the bottom of the bottle labelled B? . . . . cm^2

e) How long is the light C on for? . . . . sec

f) Do not touch it, but how hot do you think the candle flame is? . . . . °C

g) Pull the elastic bands to point E. How much force is needed? . . . . N

h) Look through the magnifying lens at F. How much does it magnify by? x . . . .

i) What is the mass (or weight) of object G? . . . . kg

Mark scheme

	narrow range 2 marks	wide range 1 mark	
a	27 – 47	17 – 57	cm
b	45 – 105	25 – 125	g
c	100 – 150	75 – 175	cm^3
d	10 – 20	8 – 30	cm^2
e	3 – 4	2.5 – 4.5	s
f	300 – 600	200 – 900	°C
g	25 – 55	20 – 60	N
h	×5 – ×10	×4 – ×12	x
i	3.5 – 6.5	2.5 – 7.5	kg

The guesses were classified as good (score 2), reasonable (score 1), and poor (score 0). The ranges for each quantity are shown in the mark scheme. Table 4.2 details the results.

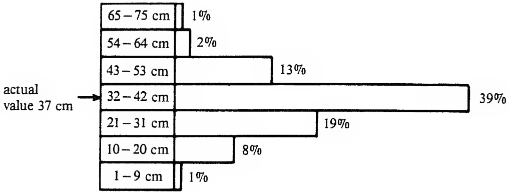
Table 4.2 *Table of performance by pupils on estimating quantities*

			% of pupils			
Quantity	Actual value	Mean score %	within narrow range (2 marks)	within wide range (1 mark)	outside wide range (0 marks)	non-response
Length	37 cm	70	64	12	24	1
Time	3.5 s	48	29	37	34	2
Area	15 cm^2	36	27	16	56	7
Magnify	×7	30	24	11	65	9
Mass (large)	5 kg	25	20	9	71	8
Force	40 N	19	13	12	75	14
Mass (small)	75 g	15	10	12	79	5
Volume	125 cm^3	12	10	4	86	13
Temp.	>200°C	6	3	7	90	5

Apart from length and time, more than half of the pupils made poor guesses concerning the values of the quantities they had been asked to estimate.

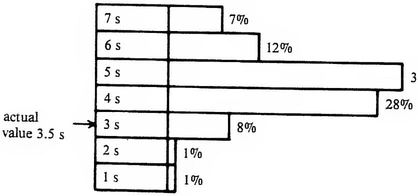
The estimates made by just over 350 pupils were used to establish the range and distribution of values given in response to this question. The results are shown in Figures 4.1 – 4.10.

Figure 4.1 *Pupils' estimates of length*



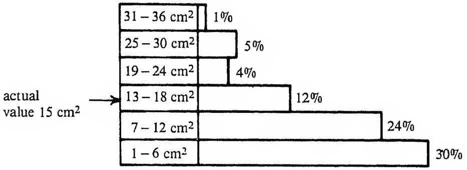
15% gave estimates over 75 cm, and 2% made no attempt. The highest and lowest estimates were 700 cm and 1 cm respectively. The majority of pupils made good estimates of the length of the object.

Figure 4.2 *Pupils' estimates of time*



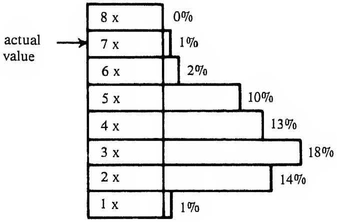
7% estimated the time interval to lie between 8 and 51 seconds. It would be reasonable to expect that the majority of 15 year old pupils have some technique for counting seconds. The shortness of the time interval and the range that was allowed may have affected the performance level as under-estimates would tend to be counted as correct responses.

Figure 4.3 *Pupils' estimates of circular area*



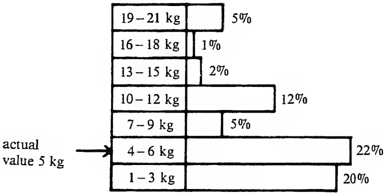
12% of the pupils gave values greater than 37 cm²: the highest value given was in fact 510 cm². The tendency to underestimate small values is quite noticeable in this graph. The majority of the guesses for both this and the ones for volume are in the ranges appropriate for the diameter of the object, possibly indicating that the pupils are making simple linear estimations.

Figure 4.4 *Pupils' estimates of magnification*



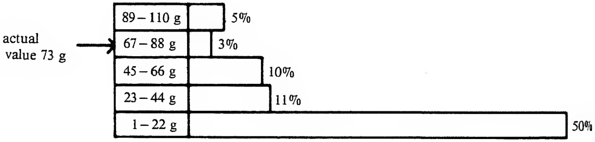
Over 20% estimated the magnification to be over ×10, some giving values as high as ×1,000. There was a high non-attempt rate. It is interesting to note that the typical pocket magnifier found in school laboratories is ×8, and that no pupil chose that value.

Figure 4.5 *Pupils' estimates of a large mass*



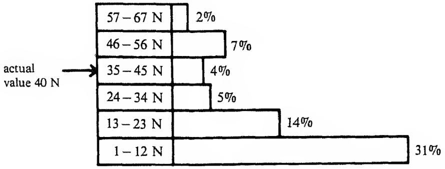
About 60% of the pupils who attempted this task overestimated the mass, which is interesting when compared with the estimate made for a small mass which is shown below. 20% of the pupils estimated the mass as being greater than 28 kg, and several thought over 2,000 kg was reasonable!

Figure 4.6 *Pupils' estimates of a small mass*



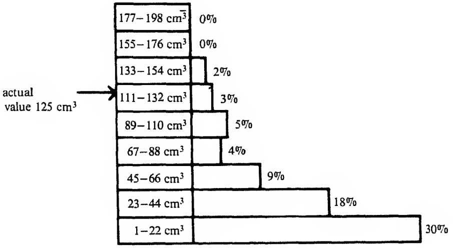
14% gave values in the range 111 – 802 g. What is most noticeable is the large percentage of pupils that underestimated the small mass, especially when compared with the previous graph. It would be interesting to know if the pupils would estimate more accurately if asked to do so in imperial units.

Figure 4.7 *Pupils' estimates of force*



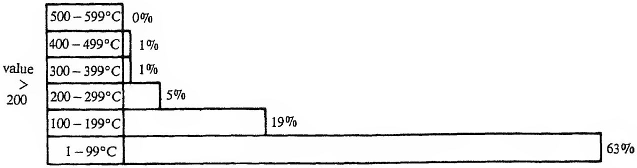
This particular quantity was found to be fairly difficult to estimate even for the teachers who were asked to validate the question, though they were usually within the "reasonable guess" range, ie. 20 – 60 N. 13% of the pupils gave estimates ranging from 68 N to 1,000 N. The high non-attempt rate was perhaps due to the relative lack of familiarity with the unit.

Figure 4.8 *Pupils' estimates of volume*



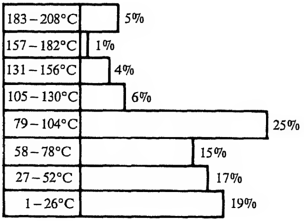
The pattern of underestimation can be seen again here. Only 6% estimated that the volume was greater than 199 cm^3, the greatest value given was 960 cm^3. It might be noted here that pupils, quite sensibly, tended to give round number estimates ie. the 5% of pupils recorded as giving values between 89 cm^3 and 110 cm^3 in fact gave choices of 100 or 110 cm^3 in the vast majority of cases.

Figure 4.9 *Pupils' estimates of candle flame temperature*



Only 1% of pupils estimated the temperature of the flame to be above 600°C. The extent of underestimating of this quantity is only fully appreciated when the values estimated by pupils are investigated in finer detail. Figure 4.10 shows the proportion of pupils that estimated in the range from 1°C to 208°C.

Figure 4.10 *Proportion of pupils estimating in the range 1°C to 208°C*



One possible explanation of this result is that the "highest" temperature that the majority of pupils are familiar with is that of boiling water and they have, therefore, scaled their estimate in relation to that "highest" temperature.

4.4.2 Summary of results on estimating

Three stations were used in this survey, and estimates were obtained on 21 quantities. The results obtained in all the questions were similar to those just described.

There would appear to be a tendency to underestimate quantities that are relatively low in value. It would be interesting to establish whether there would be a similar tendency to overestimate when the quantities to be valued are relatively large as is suggested by the results for the large mass.

4.5 Following instructions: sub-category 2γ

4.5.1 Outline and description of questions

Much practical work is dependent upon the ability of pupils to follow instructions sensibly. For example, worksheets are used to give instructions for many routine procedures such as identification of an unknown chemical by comparing its reactions with those of known chemicals.

The performance levels achieved by 15 year old pupils on questions designed to assess this ability are susceptible to wide fluctuation caused by the way the instructions are written. If the instructions only refer to a procedure rather than 'spelling out' how to carry it out the mean score drops.

The two tasks reported here both involved some recall on the part of the pupils. The first task, (Example question 6), asked the pupils to wire a three pin plug. The second task, (Example question 7), was a typical laboratory procedure presented as a worksheet.

No overall score is computed for this sub-category.

Wiring a plug

Example question 6 (2γ)

- Wire this plug correctly.
- Choose the fuse that you would put in if this plug was connected to an ordinary table lamp, and put it in.

n = 557
Mean score 34%
Non-response 2%

Mark scheme

Correct wires to correct pins	2
If this is correct then do the rest; otherwise score only for fuse	
Wire held securely in gate	2
Terminals neatly and securely held by pins, ie. no loose strands, good contact, wires well under screw.	2
fuse 3 amp	2
fuse 13 amp	1
Maximum score	⑧

raw score	% pupils
0	16
1	33
2	14
3	8
4	2
5	8
6	5
7	9
8	6

no mapping used for this question because 2γ question scores are not summed.

The type of plug that was used required all the wires to be of the same length, and the cable grip was the push-in type. The flex was cut to length and the wires were stripped ready for use. Trials had shown that few pupils had insufficient time to complete the task, if they were able to attempt it at all. 3 amp, 13 amp and 30 amp cartridge fuses were left on the desk for them to choose from.

The mark scheme used for this question made it possible to find out the parts of the task that a pupil completed correctly.

The results of the analysis are shown in Table 4.3 below. Each column shows one of the ways in which pupils scored marks. For example, 41% of the pupils scored 2 marks, at least, because they put the correct wires to the correct pins; 23% scored 4 marks at least, because they wired the plug correctly and trapped the flex in the gate. 8% scored the maximum total possible because they got all 4 parts correct. (There is a slight difference between these percentages and those shown in the raw score table on the mark scheme above because the detailed analysis of the checklist was carried out on a random sample of the total number of scripts. 135 of the 557 responses were analysed in this way.)

Table 4.3 *Breakdown of performance on the question "Wiring a plug"*

n = 135

Parts of task	% of pupils on these combinations									
Correct wires to correct pins	✓			✓	✓	✓	✓	✓	✓	✓
Wire held securely in gate						✓		✓	✓	✓
Terminals neatly and securely wired							✓	✓	✓	✓
3 amp fuse chosen		✓		✓						✓
13 amp fuse chosen			✓		✓				✓	
percentage	41	21	62	12	27	23	25	16	7	8

By studying the table it can be seen that of those who chose a 3 amp fuse a larger proportion completed the whole task correctly than those who chose the 13 amp fuse.

A few pupils experienced difficulty manipulating the parts of the plug but this was reported not to be a major problem. One administrator did report that a pupil had carefully twisted all the wires together, trapped the flex in the cable grip and laid the plaited wires carefully in the centre of the plug!

Boys scored much higher on this question than girls. (Mean scores 39%, and 23% respectively).

The second question in this group, Example question 7, was marked by a teacher from the school being surveyed who observed the pupils performing the task. A checklist was used to award marks for technique and the correct carrying out of procedures. No mark was awarded for the "answer".

The task involved 4 procedures
1. correct lighting of a Bunsen burner
2. safe heating of a solid in a test-tube over a Bunsen burner
3. non-contamination techniques when handling chemicals
4. procedure to test for a gas

1. Lighting a Bunsen burner

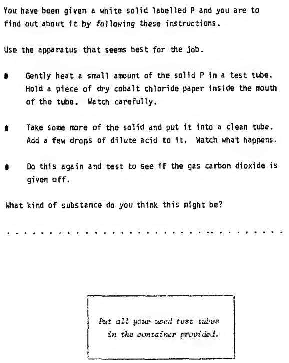
The parts of the checklist that need to be summed to find the performance level for this are a+b+c+d. The results are given in Table 4.4.

A number of schools employ spark lighters, so perhaps part c should be discounted, though the pupils were required to light the Bunsen with a match. Certainly in areas where bottle gas is used it is not safe to light the gas jet from "on top".

Following a worksheet

Example question 7 (2γ)

n = 558
Mean score 58%
Non-response 0%



You have been given a white solid labelled P and you are to find out about it by following these instructions.

Use the apparatus that seems best for the job.

- Gently heat a small amount of the solid P in a test tube. Hold a piece of dry cobalt chloride paper inside the mouth of the tube. Watch carefully.
- Take some more of the solid and put it into a clean tube. Add a few drops of dilute acid to it. Watch what happens.
- Do this again and test to see if the gas carbon dioxide is given off.

What kind of substance do you think this might be?

. . . .

Put all your used test tubes in the container provided.

Mark scheme

a. Adjusts hole to not-open	1
b. Lights match before turning on gas	1
c. Lights from below edge of chimney	1
d. Adjusts flame to non-luminous	1
e. Sensible amount of substance P used (1 – 2 spatulas)	1
f. Correct use of test tube holder	1
g. Angles test tube	1
h. Agitates test tube	1
i. Uses tweezers to handle cobalt chloride paper, ie. non-contamination	1
j. Takes a clean test tube	1
k. Non-contamination techniques used with acid, ie. does not deliberately touch sides of test tube, does not put dropper on to the bench	1
l. Chooses to use lime water	1
m. Selects a procedure for detecting gas that would work in a laboratory 1. hold drop on glass rod; 2. suck gas into teat pipette and bubble through limewater in test tube	1
Maximum score	⑬

No mapping used for this question as questions in 2γ were not summed together

It has been found that higher marks tend to be scored in questions that simply ask the pupil to light a Bunsen burner. It would appear that having this procedure embedded in a larger task makes pupils a little less circumspect in the way they go about lighting the Bunsen.

Table 4.4 *Breakdown of pupils' performances when lighting a Bunsen burner*

n = 333

	parts of task	% of pupils on these combinations								
a.	Hole turned to not-open	✓				✓	✓	✓	✓	✓
b.	Lights match before turning on gas		✓			✓	✓	✓		
c.	Lights from below lip of chimney			✓			✓	✓	✓	✓
d.	Adjust flame to non-luminous				✓			✓		✓
	percentage	51	62	43	74	34	17	13	23	18

2. Heating a solid

The parts of the checklist to be summed for this are, e + f + g + h. The results are given in Table 4.5.

Table 4.5 *Breakdown of pupils' performances when heating a solid*

n = 333

	parts of task	% of pupils on these combinations								
e.	Sensible amount of substance used	✓				✓	✓	✓	✓	✓
f.	Correct use of holder		✓			✓			✓	✓
g.	Angles test-tube			✓			✓	✓	✓	✓
h.	Agitates test-tube				✓			✓		✓
	percentage	96	82	90	37	79	89	36	74	29

Some of the techniques used by the pupils to heat the solid in the test tube suggest that they were not too familiar with the task. Several rested the test tube on the top of the Bunsen burner chimney. Some pointed the test tube at themselves, or at other people in the laboratory. These problems were also noticed by administrators running circuses where pupils had to heat chemicals as part of an observation task.

Table 4.5 clearly shows where the majority of pupils had difficulty with the procedure; they failed to agitate the test tube when heating. The relatively low score of 29% can be attributed to that factor alone.

3. Non-contamination techniques

The parts of the checklist that need to be summed for this are i + j + k, and the results are given in Table 4.6.

Table 4.6 *Breakdown of pupils' performances when using non-contamination techniques*

n = 333

	parts of task	% of pupils on these combinations						
i.	Uses tweezers	✓			✓	✓	✓	
j.	Uses clean test tubes		✓		✓	✓		✓
k.	Uses dropper pipette			✓		✓	✓	✓
	percentage	22	86	65	19	15	16	64

The cobalt chloride paper was kept in a dessicator that was clearly marked "Care. Do not let paper get damp before use".

The 22% of pupils that did use tweezers to handle the cobalt chloride paper had in their number some who did not use a clean test-tube for the different tests. 15% of pupils did consistently employ non-contamination techniques.

4. Testing for a gas

The parts of the checklist that need to be summed for this are l + m, and the results are given in Table 4.7.

Table 4.7 *Breakdown of pupils' performances when testing for a gas*

n = 333

	parts of task	% of pupils on these combinations		
l.	Chose limewater	✓		✓
m.	Appropriate procedure		✓	✓
	percentage	27	8	8

The pupils had to choose from cobalt chloride paper, litmus paper, water or limewater: the number of pupils who chose limewater is possibly a little higher than it would have been had a larger range of alternatives been offered.

Several novel ways of testing for a gas were encountered! A fairly common one was for the pupil to put some liquid, eg. limewater, into a clean test tube and then invert this over the test tube with the hot solid in. Other ways included attempting to capture any gas by inverting an empty test tube over the hot solid, then putting this "gas-filled" test tube into a rack and adding limewater.

Out of the whole sample of 558 pupils tested on this question, one got all of the procedures marked correct, and five got 12 of the 13 points.

There was no sex difference in the mean scores for this question.

4.5.2 Summary of results for following instructions

These questions were the only two used in the survey in this sub-category. In future years it is hoped to include others that focus on different procedures.

4.6 Summary of Category 2

The practical skills assessed in this survey ranged from a knowledge of units appropriate to the instrument, through the reading of scales and the use of simple measuring apparatus, to well defined and fairly complex procedures for measuring small quantities. Order-of-magnitude estimates of common quantities and two standard laboratory exercises were also assessed.

It was found that the type of scale, particularly the way it is subdivided, has a significant effect upon the level of performance. When the value had to be read from a linear scale marked in single major divisions and the reading lay on the division, then between 60 – 80% of the pupils succeeded. As the complexity increased the performance dropped to about 30%. Kellington and Mitchell (1979) have reported similar results for younger Scottish pupils.

Low performance levels were also found on the use of certain measurement procedures. In particular, measurement of small distances and short time periods gave mean scores below 20%, indicating that few pupils employ an appropriate technique.

The majority of pupils were able to give reasonable estimates for linear dimensions and for time whereas estimates for area, force, volume, and particularly, temperature were less accurate.

Swan and Jones (1980) report the results of estimation exercises carried out in 1971 and 1977 which were similar in content to the questions used in this survey, but with very different magnitudes. Their subjects were better at guessing length than weight or temperature and appeared to be making linear estimates of area and volume — results which broadly support our findings.

Using the 1980 survey data for Category 2 it is not possible to draw well-formed general conclusions about the performance levels of pupils on practical skills. They do, however, suggest that knowledge of some of the basic practical skills is not as widespread as might have been anticipated.

5 Using observation

5.1 Introduction

Observation plays a crucial role in science. Many science curricula and examination board syllabuses emphasise in their aims and objectives the need to develop observation skills in pupils. These skills also feature prominently in the criteria used in HMI's Secondary Survey (DES 1979). The importance of observation in science has led to a considerable amount of philosophical reflection on what the act of observing involves.

In the context of science, observation is not restricted to the sense of sight. It includes all the senses, and it is not a passive activity. Merely "seeing" something could be adequately tested by employing the methods used by opticians. When engaged in scientific observation the observer uses an intellectual framework in order to know what to see; how to look at the world. This framework is acquired through experience and training in the discipline of science. Ideas and theoretical notions about the world lead to the creation of a model which informs about what to select as noteworthy: what to disregard as spurious. Unfamiliar objects and events are interpreted in the light of what is new compared to what is already accepted. This process of analogy is fundamental to the use of models. The possession and use of the particular types of model found in science distinguishes scientific from other types of observation, such as artistic "appreciation" guided by aesthetic principles. It is considerations such as these that have led to the current view that scientific observation is "theory-laden".

In this category the questions and the way they are marked at age 15 reflect this theory-laden view. A variety of question types are employed to assess the extent to which pupils use some form of scientific framework in making and interpreting observations. It is impossible to isolate the act of communicating the observations made from the act of observation itself. For this reason questions also vary in the methods by which pupils are required to record their observations.

Questions in Category 3 are grouped into three sub-categories:

sub-category 3α	Using a branching key — pupils are required to identify objects using a branching key.
sub-category 3β	Observing similarities and differences — pupils are required to make comparisons either between two or more objects or events, or between successive stages of an event. They may have to decide what is important and relevant in the comparison and record their decisions in a variety of ways.

sub-category 3γ Interpreting observations
— pupils are required to collect information by observation often by making comparisons or noting changes, and then use what they take to be relevant to go beyond their observations to find patterns, explain what they observe, or predict.

In what follows each sub-category is discussed in turn, with the help of some illustrative questions. A short summary will be found at the end of each section indicating the results and any general conclusions that can be drawn. The first section briefly outlines the way in which the observation tasks were administered and the resources that were used.

5.2 Administration

Questions used to assess observation skills in a large survey such as this require stimulus materials which are very reliable and are of a high standard.

Equipment required for the tests was made to very detailed specifications, and where photographs were used these were often commissioned specially. Trained administrators were used to take the test materials to the schools and to administer the tests. The "circus" format outlined in Chapter 4 was used. (More detailed information about the way in which the tests were administered can be found in Appendix 3.)

It is not possible in a publication of this kind to reproduce faithfully the quality of photographs that the pupils saw or to give more than an indication of the apparatus used. Readers are asked to bear this in mind when viewing the questions which follow.

5.3 Using a branching key: sub-category 3α

5.3.1 Outline and description of questions

Questions that require pupils to use a branching or binomial key to identify objects are grouped separately from other questions in Category 3. This is because the demands made upon the pupils by the reading and the logical structure of the key may well contribute at least as much to the pupils performance as the directed observation. At age 15 this is the only type of question which the administrator is not allowed to read to a pupil who has reading difficulties.

All of the questions in this sub-category ask the pupils to name three objects using a key. Unfamiliar names are deliberately chosen to reduce the possibility of a pupil naming an object correctly from memory.

Three questions were used in this year's survey. Each one was the second of two questions at a six minute station. The first question at each station was a short multiple-choice question leaving the pupils about five minutes to attempt the key question.

Example question 1 is typical of those set. Technical words were avoided, or, as in this case, the meaning was illustrated by a drawing. The three grasses that had to be identified were mounted on card and were set out so that they could be viewed easily—all the salient features were clearly visible.

The objects to be identified in the other two questions of this type used this time were three seaweeds and three worms.

Two common types of error made by the pupils were: listing the number of the sentences they read without giving a name to complete the question; and responding to each sentence in the key with either a "yes" or a "no", again without attempting to name the object. It was noted that a considerable number of pupils did not copy the names from the key accurately. They were marked as correct whenever their intent was clear.

Identifying objects using a key

Example question 1 (3α)

n = 558
Mean score 43%
Non-response 3%

Look carefully at the three grasses A, B and C and use the identification key below to name them.

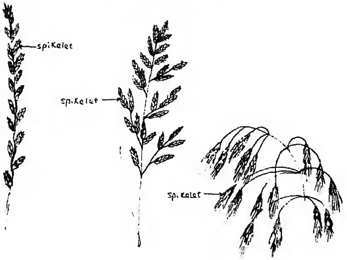
1.	Spikelets borne on a single stem	If yes go to 2
	Spikelets borne on branches of stem	If yes go to 3
2.	Spikelets so close together that they are touching and together look like a single structure	If yes go to 4
	Spikelets further apart	If yes go to 5
3.	Branches bearing spikelets arise from stem singly or in pairs and each branch may have several spikelets	Meadow fescue
	Branches arise from stem in groups of three or four and each branch has only one spikelet	If yes go to 6
4.	Spikelets form a dense cylindrical spike which is rounded at each end	Timothy grass
	Spikelets form a dense cylindrical spike which has a point at each end	Meadow fox-tail
5.	Spikelets arranged alternately on stem	Perennial rye-grass
	Spikelets are closer together	Couch grass
6.	Spikelets borne on short drooping branches Lower branches are longer than upper ones giving a tapering shape	Smooth meadow grass
	Spikelets borne on long curved drooping branches and have a delicate hairy appearance	Barren brome

A . . . .

B . . . .

C . . . .

Examples of some grasses



Mark scheme

A.	Perennial rye-grass	1
B.	Meadow fescue	1
C.	Timothy grass	1

Maximum score ③

mapped score	% pupils
0	34
1	25
2	21
3	21

5.3.2 Summary of performance on sub-category 3α

The mean scores were 25% for the "worm" key, 43% for the "grass" key and 46% for the "seaweed" key. (The low 25% was due to a high proportion of non-attempts — 20%). These results, particularly when taken together with information obtained from large scale trials, tend to suggest that the level of performance in using a binomial key may be relatively unaffected by the type of object to be identified.

5.4 Observing similarities and differences: sub-category 3β

5.4.1 Outline and description of questions

Observation plays a part in school science activities in a range of ways. Typical types of observation tasks have been identified and form the basis of the questions in this sub-category. (Nine types of tasks are used in the tests and these are listed in Appendix 1.) The types of task include noting similarities and differences between specimens given to the pupils, classifying objects using observable features, recording the changes over time that occur in an event, and making a scientific drawing. In all cases the questions require pupils to make comparisons based on observable features: either between two or more objects, between an object and a representation of it, or between observations made at different times. In any observation task there is a considerable number of possible observations that can be made, many of them redundant. In order to make the required comparisons pupils need to select those aspects which they consider to be relevant.

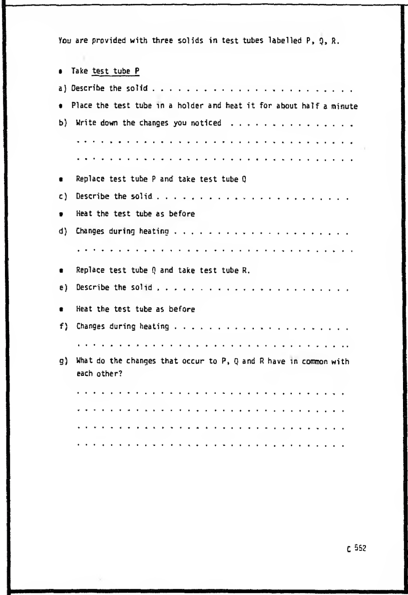
Four questions are used to illustrate the sub-category. Two needed a written response, and two a drawing.

Example question 2 required pupils to describe a chemical event. They were given three test-tubes: the first contained crystals of iron (II) sulphate, the second crystals of copper (II) sulphate and the third crystals of nickel (II) sulphate. The pupils were asked to describe each substance in turn and to say what happened to the contents of the tubes when they were heated. The first six parts of the question required fairly low-level observations to be made, concerned with such aspects as colour changes and the description of evolved substances. The last part of the question is an attempt to see whether pupils can generalise about their observations, ie. whether they can identify the similarities from amongst the differences.

It was generally found that the colour descriptions were done well. The next most popular observation was the vapour condensing on the tube. Identifying the similarities was done poorly.

Observing an event

Example question 2 (3β)



You are provided with three solids in test tubes labelled P, Q, R.

- Take test tube P

a) Describe the solid . . . .

- Place the test tube in a holder and heat it for about half a minute

b) Write down the changes you noticed . . . .

. . . .

. . . .

- Replace test tube P and take test tube Q

c) Describe the solid . . . .

- Heat the test tube as before

d) Changes during heating . . . .

. . . .

- Replace test tube Q and take test tube R.

e) Describe the solid . . . .

- Heat the test tube as before

f) Changes during heating . . . .

. . . .

g) What do the changes that occur to P, Q and R have in common with each other?

. . . .

. . . .

. . . .

. . . .

C 552

n = 756
Mean score 49%
Non-response 0%

Mark scheme

a.	Blue, turquoise, colour (light, pale, green)	1	
	crystals (no mark for powder)	1	2
b.	Brown or white solid	1	
	noise produced OR vapour produced	1	any up to
	vapour condenses or changes into	1	3
	clear liquid (water) formed	1	
	powder formed	1	
c.	Blue colour	1	
	crystals	1	2
d.	Turns white	1	
	then turns black or yellow/brown	1	
	vapour or steam or gas produced	1	
	(changes to) clear liquid (water)	1	4
e.	Green colour	1	
	crystals	1	2
f.	Vapour given off	1	
	clear liquid (water) formed	1	
	yellow solid formed	1	3
g.	All give off vapour/water		
	All change colour any *one*	1	1

Maximum score ⑰

mapped score	raw score	% pupils
0	0, 1, 2, 3	11
1	4, 5, 6, 7, 8	38
2	9, 10, 11, 12, 13	44
3	14, 15, 16, 17	7

Pupils doing this question were observed by a teacher from the school. The teacher was instructed to help any pupil who had difficulty with the experimental procedure. In this way failure to score could confidently be attributed to either faulty observation or poor recording of those observations. Several administrators reported that the teachers met with a fairly high proportion of pupils who required some help with the technique of heating solids in a test-tube over a Bunsen flame. (This is in keeping with the results reported for Example question 7 described in Chapter 4.)

Example question 3 offered the pupils three black and white photographs of different, but related, beetles. They needed to compare and contrast the three in order to discriminate between them. No knowledge of special terms was necessary, but the characteristics chosen to base the comparison on had to be related to the structure of the beetles and not likely to be merely artefacts caused by the production of the photograph.

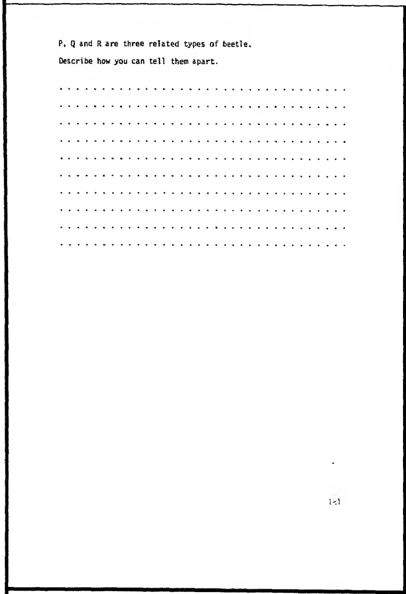
Several of the responses simply repeated the question, stating that the beetles were different, but not indicating how they differed. Others tended to speculate about where or how the beetles lived. A fairly common error was the repetition of a particular point, for example stating that "P is bigger than Q" and "Q is smaller than P".

In the marking of this question it was noticed that the answers which indicated an orderly scan of the beetles, starting from the head and working downwards, or vice versa, tended to score more highly than those that adopted an alternative strategy.

Comparing and contrasting

Example question 3 (3β)

n = 773
Mean score 49%
Non-response 0%



P, Q and R are three related types of beetle.

Describe how you can tell them apart.

. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .
. . . .

mapped score	raw score	% pupils
0	0, 1	14
1	2, 3	38
2	4, 5	36
3	6	12

Mark scheme

Common comparisons made shown below
Others are acceptable if justified

P has any 2 of		
biggest	1	
blacker	1	
more dots on back	1	
eyes set back	1	
large (crushing) (open) mandibles/nippers/claws	1	
thorax shape different	1	
hairs on the "nose" different number	1	
antennae start immediately behind mandibles (well forward)	1	
Any two		2
Q has any 2 of		
fatter antennae	1	
crossed over mandibles/nippers/claws	1	
palps (. . . . . . .) set well back	1	
fewest holes on back/widely spaced	1	
hairs on "nose" differently arranged	1	
first pair of legs set well forward	1	
indentation in the abdomen/body	1	
egg-shaped abdomen/body	1	
Any two		2
R has any 2 of		
most hairy	1	
smallest	1	
mandibles touch/are pointed	1	
antennae start near eyes	1	
largest holes in the back	1	
most pronounced waist	1	
palps (. . . . .) squarish	1	
first pair of legs set well back	1	
Any two		2
If characteristics are given but not ascribed score 1 out of 2	1	
Maximum score		⑥

In the third question in this section, (Example question 4) the pupils were given two photographs, A and B. A was of a root section, taken at low power, and B was part of the vascular bundle of the root section at a higher power. The task set to the pupils was to identify which part of A had been magnified.

The results show that about half of the pupils located the general area that had been enlarged. The majority of these pupils, however, failed to make the detailed discriminations necessary to identify the exact location.

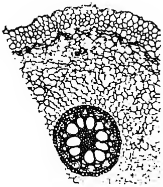
Identifying a part in a whole

Example question 4 (3*β*)

Photograph A is of a root cut across (x 40).

Photograph B is a part of the same root, but under much higher magnification (x 400).

Look at the diagram below and put a rectangle around the exact region which you think is shown in photograph B.

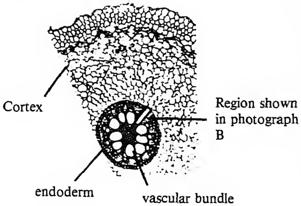


C 177

n = 756
Mean score 29%
Non-response 8%

Mark scheme

Exact region (ie. should include endoderm)	3
Too large, too small, but exact region	2
Includes endoderm and vascular bundle but wrong area	1
Maximum score	③



mapped score	% pupils
0	47
1	28
2	16
3	9

Example question 5 asked the pupils to make a "scientific" drawing of a lobster. The pupils were given a specimen of a lobster mounted, on its side, in a box.

This type of question in effect is similar to Example question 2 in that the pupils are asked to make a report of their observations. In this case, however, they were asked to do it pictorially. The difference between a "scientific" drawing and any other type of drawing is reflected in the mark scheme. A checklist was drawn up of the significant features that should be included in the drawing. No reference to the artistic merit of the drawing was made. For example, the number of legs, the point of attachment, the joints and the claws at the end of the legs needed to be shown, but the actual arrangement of the legs was not significant. Thus it could be said that what was required was a diagrammatic record rather than a representational drawing.

Nearly half of the pupils made drawings which incorporated over half of the features identified on the mark scheme. The non-response rate was very low.

Two other drawing questions were used in this 1980 survey. One was of a locust, the other was of a section through a stem. The scores for these questions are shown in Table 5.1, the results for the "lobster" question being repeated for convenience.

Table 5.1 *Score distributions for questions involving drawing*

mapped score	*Percentage of pupils on questions called* "lobster"	"locust"	"stem"
0	14	13	26
1	41	54	47
2	32	25	20
3	13	8	7

The mapped score points for each question reflect approximately the same number of features noted in the drawing.

It can be seen that the levels of performance are similar for the three questions.

Make a scientific drawing

Example question 5 (3β)

This is a lobster.

Look at it from the side and make a scientific drawing of the animal in the space below.

Your drawing should be in pencil. It should be as large and as clear as possible.

185

n = 773
Mean score 48%
Non-response 1%

Mark scheme

Four legs from thorax — must come from separate points	1
Legs jointed	1
Legs end in claws	1
Three parts to claw (major)	1
Claw (major) proportioned: (nippers — ⅓ claw)	1
Two pairs of antennae	1
Palps	1
Head correct (eye position (general) rostrom)	1
Tail fan	1
Tail fan segmented	1
Abdomen segmented	1
Five or six elements to abdomen	1
General proportion correct	1
Size of drawing adequate (fills about ¾ of the page)	1
Maximum score	⑭

No half marks to be given

mapped score	raw score	% pupils
0	0, 1, 2, 3, 4	14
1	5, 6, 7, 8	41
2	9, 10, 11	32
3	12, 13, 14	13

5.4.2 Summary of performance on sub-category 3β

17 questions in this sub-category were used in the 1980 survey. The mean score was 43%. The maximum score achieved was 80% and was attained by 4 pupils out of the sample of 770. 3% of the pupils failed to score at all on the test.

It is recognised that some pupils record low marks because their skill in communicating what they have observed is limited.

5.5 Interpreting observations: sub-category 3γ

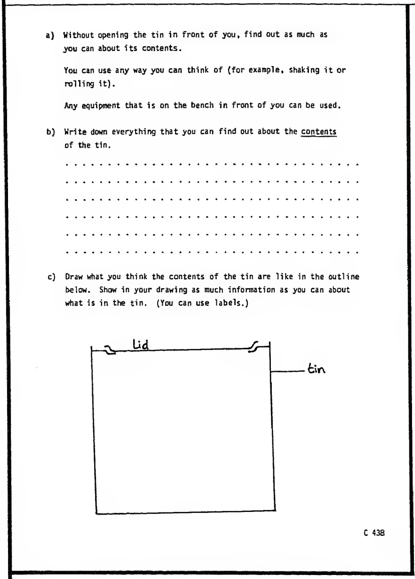
5.5.1 Outline and description of questions

In this sub-category, the questions require the pupils to collect information by observation, and then, using what they take to be relevant, to explain their observations, or to make predictions based on them.

9 questions from this sub-category were included in the survey. Two of the questions are discussed in detail on the following pages.

Going beyond the observations

Example question 6 (3y)



a) Without opening the tin in front of you, find out as much as you can about its contents.

You can use any way you can think of (for example, shaking it or rolling it).

Any equipment that is on the bench in front of you can be used.

b) Write down everything that you can find out about the contents of the tin.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

c) Draw what you think the contents of the tin are like in the outline below. Show in your drawing as much information as you can about what is in the tin. (You can use labels.)

C 438

n = 756
Mean score 36%
Non-response 0%

Mark scheme

a. *Observations*

Any to a maximum of 8
Does not roll
Tin lopsided *or* weight fixed on one side
(no shaking around)
Heavy
Seems magnetic (NOT "is magnetic")
Not a magnet
Prefers or stops on one side
Sounds hollow on one side
Others as suitable (eg. no liquid in tin) 8

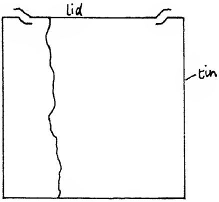
b. *Inference*

A.	Drawing is consistent with all of their observation and mass of no specific shape, and does not contradict any of the above	4
B.	Drawing that does coincide with their observations and contradicts only one of the above	2
C.	Drawing that contains totally speculative elements, eg. colour, and is otherwise as A	3
D.	Drawing contains speculative elements and is otherwise as B	1
E.	Drawing is not consistent with all their observations, but is in fact correct	2
F.	No observations and no drawing	0
G.	Observations and no drawing	0
N.	No attempt Not categorised	4

Maximum score ⑫

Shape should be generalised, but symmetrical, top to bottom

Side view



raw score	% pupils
0	15
1–3	28
4–6	31
7–9	18
10–12	7

Example question 6 is one of a series of "black box" questions. In these questions pupils were provided with a sealed tin, containing a simple device (such as a lump of plasticine fixed to one side), a ruler and a magnet. Pupils had to make observations about the contents of the tin without opening it. They then had to draw what they thought was in the tin.

Both the observations and the quality of the inferences they drew from them were marked. The inferences pupils made about the contents were judged in terms of whether they accounted for the observations they had made.

About a quarter of the pupils' drawings would have accounted for all the observations they made and another 60% made drawings based to some extent on the observations they had recorded.

It has been found from trials of a number of these questions that the categories of response remain fairly stable even though the contents of the tin are changed. Table 5.2 shows the proportion of pupils that fall into the categories outlined in the mark scheme.

Table 5.2 *Percentage of pupils in each category*

Category	*%*	*Meaning of category*
A	25	drawing consistent with observations of pupil; does not contradict mark scheme
B	16	drawing consistent with observations of pupil; contradicts one on mark scheme
C	7	drawing consistent with observations of pupil; does not contradict mark scheme, but contains a speculative element eg. colour
D	15	drawing consistent with observations of pupil; contradicts one listed on mark scheme, and contains a speculative element
E	1	drawing correct, contradicts observations made by pupil
F	13	drawing wrong, no observations recorded
G	17	no drawing, pupil's observations correct
N	1	No attempt recorded on question paper
*	5	Responses not categorised

The second question used to illustrate this sub-category is much more concept-laden. In Example question 7 the level of performance could be expected to be affected by pupils' knowledge of physics. The pupils had four wires mounted on carriers labelled A to D and of decreasing thickness. They could put wires A to C into the circuit, and had to predict what would happen to wire D if it was put into the circuit.

The observation of the wire is interesting. The movement due to expansion caused by the heating effect is clearest in the case of wire C and is noticeable for wire B when this is being looked for. This suggests that a good strategy would be to try the wires A, B, C and then to check the observations by rerunning the trials. 85% of the pupils scored on the observation part of the question. Most, however, (about 65%) recorded the current changes and did not comment on the alteration in the state of the wires B and C. In the prediction part, though, 36% referred to the state of the wire D and described adequately what would happen to it.

In order to gather some information about what the pupils did when faced with the apparatus a teacher from each survey school was asked to observe the pupils as they attempted this question and to fill in a checklist.

For each of the sub-groups in the four checklist categories the following guidelines were given to the teachers:

confidence: The teachers were asked to identify the degree of confidence the pupil showed. As can be seen, in general, the teachers' perception of this behaviour in pupils conforms to the expected distribution.

Good	29%
Average	47%
Poor	24%

level of interest: Three levels were defined:

none	— pupil evinces no interest in the apparatus and engages in the minimum amount of activity necessary to answer the question	18%
curious	— pupil examines the apparatus visually without attempting to handle it	57%
active	— pupil actively inspects apparatus.	25%

repetition: The observing teacher was asked to note the number of trials the pupils made before recording their conclusions.

0	repetitions	53%
1 or 2	repetitions	32%
3+	repetitions	15%

time to form conclusions: Timing began from the instant the pupils, having read the question, began to use the apparatus until they completed their prediction.

Up to 2 minutes	46%
2 to 5 minutes	36%
5 to end of station	11%
No conclusion reached	5%
Not classified	2%

In general, it was found that pupils who lacked confidence when handling the apparatus were less likely to show interest, less likely to repeat any observation and more likely to have finished within five minutes than pupils identified as being confident.

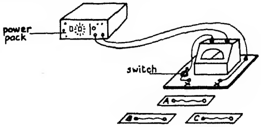
It is interesting to note the relatively few pupils who attempted any retrials and the speed with which many came to a conclusion. This, presumably, indicated they were satisfied they had accumulated sufficient information to make their prediction.

Other questions in this sub-category required pupils to use other science concepts and knowledge to interpret the observations they had made.

Explaining your observations

Example question 7 (3γ)

In front of you is the circuit drawn below:



The circuit can be completed by plugging in the apparatus labelled A, B or C at the points labelled X, and then turning on the switch.

- Plug A into the circuit.
- Switch on.
  Look carefully at the wire and then the meter.
- Switch off.
  Write down what you saw.

- Now plug in B.
- Switch on.
  Look carefully at the wire and then the meter.
- Switch off.
  Write down what you saw.

- Plug in C.
- Switch on.
  Look carefully at the wire and then the meter.
- Switch off.
  Write down what you saw.

	What happened to the wire	What happened to the meter
A	. . . .	. . . .
B	. . . .	. . . .
C	. . . .	. . . .

- Look at D.
  You cannot put this into the circuit, but if you could, what do you think would happen to the wire and to the meter?

  . . . .

  . . . .

  . . . .

  . . . .

n = 773
Mean score 47%
Non-response 2%

Mark scheme

Current lessens A→C	1	
or wire sags or gets hotter A→C	1	
		1
wire		
gets very hot, or melts	2	
or wire will sag (expand)	1	
meter		
current less than lowest reading (meter moves very little)	1	
		3

Maximum score ④

raw score	% pupils
0	15
1	26
2	23
3	27
4	9

5.5.2 Summary of performance on sub-category 3γ

The mean scores obtained for the nine questions set in this survey varied from 32% to 57%. There is an indication that pupils' levels of performance on questions of this type depend, as might be expected, on their familiarity with the scientific ideas involved.

5.6 Summary of Category 3

The types of questions that make up this category all require the pupils to be actively involved in the process of observing. Various other elements, such as reporting, comparing and inferring, all contribute to the final level of performance measured. The results for 3β, for instance, do not mean that on average pupils only see about 40% of what is happening around them. They do suggest, however, that observing the world in the way that science requires is a skill that needs to be acquired.

By the time pupils are in the fifth form it may be expected that they will have been introduced to a scientific view of the world. This exposure offers the pupils another way of looking at objects and events and guides their selection of "relevant" observations. The tasks were designed so that pupils could score whenever they made a valid observation about the object. In general it was found that about half the pupils got about two thirds of the marks allocated to the observation tasks and that most pupils could get at least one third.

One of the most common errors made by pupils was the inclusion, as an observation, of information they could recall about the object but which was not observable. For example some claimed that wasps and ants differed from flies because both of the former could sting, or that chemicals were alike because they were all made up of atoms.

It is possibly worth noting here, though it will be discussed more fully in Chapter 9, that pupils studying biology had a higher level of performance on observation tasks than pupils who study no biology at this age.

6 Interpretation of data

6.1 Introduction

Category 4 is divided into two parts. One part is concerned with the ability to interpret data; the other with the ability to recall and apply science concepts in given situations. In this chapter we describe the questions and results relating to the first of these, ie. interpretation of data. Pupils performances on application of science concepts are discussed in the following chapter.

The process of looking for patterns or relationships in data plays an important part in many activities including science. Data from experiments or observation are reviewed for regularities. In some cases apparently anomalous results can be identified and checked and further predictions can be made on the basis of perceived relationships. In other cases, the data is inspected to assess the extent to which it supports a particular hypothesis.

The assessment framework reflects the types of activity outlined above. There are three sub-categories concerned with the interpretation of data:

4α Describing and using patterns in information.

4β Judging the applicability of a given generalisation.

4γ Distinguishing degrees of inference.

A section of the chapter is devoted to each of the three sub-categories. In each section we illustrate the nature of the sub-category with example questions, before discussing overall results.

6.2 Describing and using patterns in information: sub-category 4α

Questions in this sub-category assess the extent to which pupils discern patterns or relationships in data when all the necessary information for an answer is provided in the question. In some questions the information is presented in pictures and diagrams; in other questions where the information is quantitative in nature, it is presented in tables or graphs.

One way of assessing whether pupils discern patterns in data is to ask for a description of the relationship or pattern in presented data. Another possibility is to assess whether pupils can use the pattern in the data to make predictions. Both approaches were used in the questions in the survey.

The survey included 51 questions in this sub-category, with 17 in each of three test booklets. Examples of four different types of question together with performance data are presented and discussed in the next section.

Making predictions based on patterns in data

In questions of this type, pupils were presented with data and they had to use it either to make a prediction or to give a short explanation based on the pattern of the data.

In Example question 1, pupils were given the results of two experiments in a pictorial form, and they were to make a prediction based on the relationship inferred from these drawings. There was a high level of performance on this question with the mean score for boys higher than that for girls (88% compared with 79%). There was a lower level of performance on Example question 2, "carbon cycle". In this question pupils had to use information presented in a flow chart. The types of response pupils gave were categorised and the proportion of pupils giving each type of response is given with the question. About a third of the pupils used the data in the question to give an adequate response.

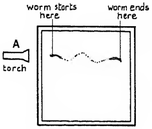
Ten questions of this type were used in the survey. Their mean scores ranged from 30% to 85%. Questions requiring a short numerical answer, or an answer in a graphical form, in general had a higher performance level than questions requiring a prose answer. A higher non-response rate for questions requiring a prose answer could have contributed to the lower overall performance on these questions.

"Worm"

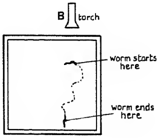
Example question 1 (4*a*)

A boy shone a torch on a worm and watched where the worm moved.

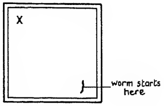
This is what happened when he put the torch at A.



He then moved the torch to position B.
This is what happened.



Draw on this diagram where you would put the torch to get the worm to position X.

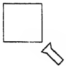


C 756

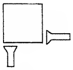
n = 773
Mean score 84%
Non-response 1%

Mark scheme

EITHER torch placed diagonally



OR in successive positions

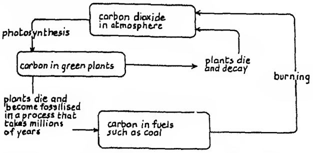


Maximum score ①

"Carbon cycle"

Example question 2 (4*a*)

The diagram below shows part of the carbon cycle:


carbon dioxide in atmosphere
photosynthesis
carbon in green plants
plants die and decay
burning
plants die and become fossilised in a process that takes millions of years
carbon in fuels such as coal

It is said that supplies of coal will run out in [illegible] years.

Use the information above to explain why the coal will run out.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

C 332

n = 739
Mean score 31%
Non-response 13%

Mark scheme

Any statement indicating differential rates of coal formation and use

Maximum score ①

Categories of response	% pupils
Differential rates of coal formation and use	31
Slow rate of formation of coal (no mention of rate of use)	13
Shortage of requirements to make coal (eg. CO_2 in atmosphere running out)	29
Confused responses	12
Tautology	1
No attempt	13

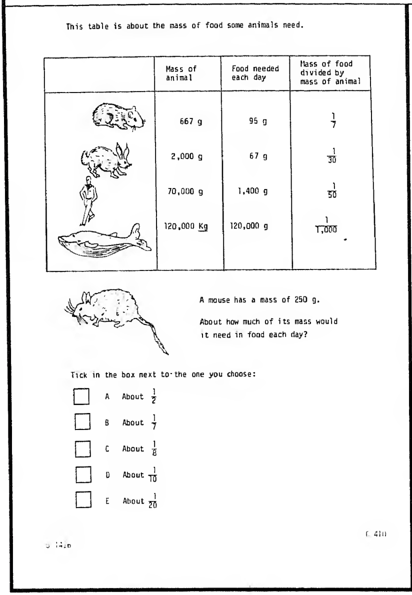
Selecting predictions based on patterns in data

Questions of this type are similar to those just described. However, instead of making a prediction, in this case pupils have to select the correct prediction from a list of alternatives. There was no overall difference in level of performance between the coded answer and the free response form of these questions.

Example question 3 is one of nine such questions used in the survey. In this question pupils have to relate three variables in order to work out an answer. The mean score for the question was about 50%. The same question was also included in the 1980 science survey of 13 year old pupils and the results for the two ages are given for comparison. Mean scores for other questions of this type ranged from 35% to 80% with most having a mean score of around 50%.

"Food needs"

Example question 3 (4*α*)



This table is about the mass of food some animals need.

	Mass of animal	Food needed each day	Mass of food divided by mass of animal
	667 g	95 g	$\frac{1}{7}$
	2,000 g	67 g	$\frac{1}{30}$
	70,000 g	1,400 g	$\frac{1}{50}$
	120,000 Kg	120,000 g	$\frac{1}{1,000}$

A mouse has a mass of 250 g.

About how much of its mass would it need in food each day?

Tick in the box next to the one you choose:

☐ A About $\frac{1}{2}$

☐ B About $\frac{1}{7}$

☐ C About $\frac{1}{8}$

☐ D About $\frac{1}{10}$

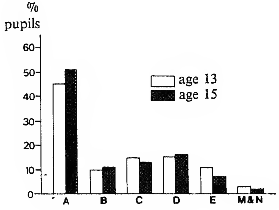
☐ E About $\frac{1}{20}$

C 410

S 142b

n = 739
Mean score 51%
Non-response 2%

Option	% pupils
*A	51
B	11
C	13
D	16
E	7
Non-response	2



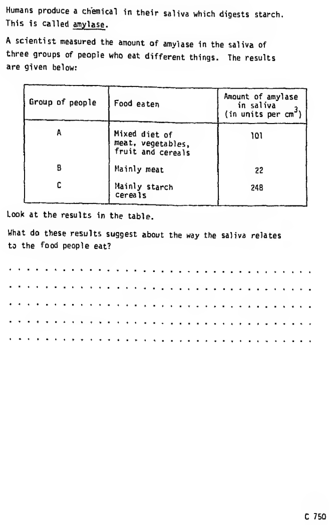
Describing patterns in data

A number of questions presented data relating two variables in tabular or graphical form. Pupils were asked to describe the pattern or relationship. Example question 4, "saliva", is of this type. Three other similar questions were also used in the survey. About a third of the pupils identified the variables and described the relationship between them in a general way.

In questions where the pattern was qualitative the level of response was higher with about two thirds of the pupils correctly describing the relationship. There was a consistent non-response rate of about 10% for this type of question; this was higher than for any other type of question in the sub-category.

"Saliva"

Example question 4 (4*a*)



Humans produce a chemical in their saliva which digests starch. This is called amylase.

A scientist measured the amount of amylase in the saliva of three groups of people who eat different things. The results are given below:

Group of people	Food eaten	Amount of amylase in saliva (in units per cm^3)
A	Mixed diet of meat, vegetables, fruit and cereals	101
B	Mainly meat	22
C	Mainly starch cereals	248

Look at the results in the table.

What do these results suggest about the way the saliva relates to the food people eat?

. . . .

. . . .

. . . .

. . . .

. . . .

C 750

n = 739
Mean score 38%
Non-response 6%

Response	% pupils
The more starchy foods eaten, the more amylase in the saliva	22
Statements indicating amount of amylase produced depends on type of food eaten	39
Other	42
Non response	6

Some questions asked pupils to describe a pattern and to use it to make or select a prediction. In some cases the pupils were required to make the prediction first and then describe the pattern on which it was based. Overall the performance level did not appear to be affected by whether it was the prediction or pattern which was asked for first. The level of performance on this type of question was lower than for any other question types in this subcategory. The mean scores ranged from 27% to 61% of the maximum scores, with around 25% of pupils correctly stating the pattern and making a prediction.

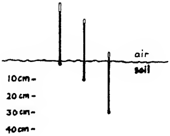
Example questions 5 and 6 are of this type. The latter question on "planets" was also set to 11 and 13 year old pupils. The types of response given at each of the three ages were categorised and the results are given with the question. As these results indicate, nearly half the 15 year old pupils (as compared with about a fifth of the 11 year olds) correctly described the relationship in the data. About 20% gave either an incomplete or incorrect description and the remaining pupils responded by making irrelevant or incorrect references to parts of the data.

"Soil temperature"

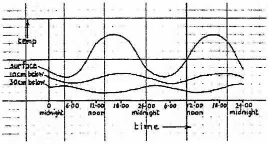
Example question 5 (4α)

Some scientists did an experiment to find the temperature of soil at three different depths:
- at the surface,
- 10 cm below the surface, and
- 30 cm below the surface.

They measured the temperature of the soil regularly over 48 hours.



The temperatures at three depths are shown in the graph below:



a) State two things the graph shows about the way the temperature variation depends on the depth of the soil.

1 . . . .

. . . .

2 . . . .

. . . .

b) About what time would the soil reach its maximum temperature at a depth of 20 cm? (Tick one box.)

☐ A 14.00 hr
☐ B 16.00 hr
☐ C 18.00 hr
☐ D 22.00 hr
☐ E 2.00 hr

434

n =	761
Mean score	30%
Non-response	6%

Mark scheme

a. (i) Maxima occur at different times 1

(ii) Difference between maxima and minimum varies with depth 1

b. Key D 1

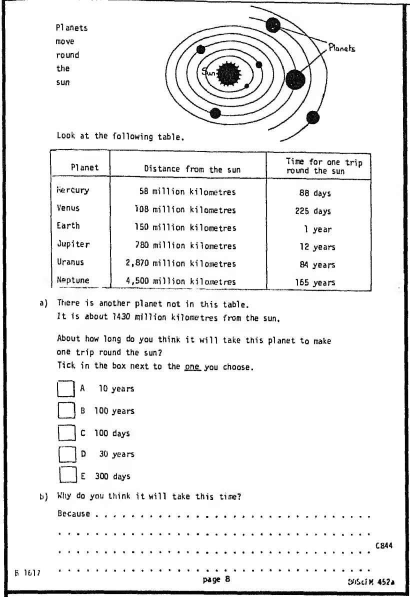
Part b

Option	% pupils
A	10
B	14
C	36
*D	28
E	6

Score for part a and b	% pupils
3	6
2	18
1	38
0	39

"Planets"

Example question 6 (4α)


Sun
Planets

Planets move round the sun

Look at the following table.

Planet	Distance from the sun	Time for one trip round the sun
Mercury	58 million kilometres	88 days
Venus	108 million kilometres	225 days
Earth	150 million kilometres	1 year
Jupiter	780 million kilometres	12 years
Uranus	2,870 million kilometres	84 years
Neptune	4,500 million kilometres	165 years

a) There is another planet not in this table.
It is about 1430 million kilometres from the sun.

About how long do you think it will take this planet to make one trip round the sun?
Tick in the box next to the one you choose.

☐ A 10 years
☐ B 100 years
☐ C 100 days
☐ D 30 years
☐ E 300 days

b) Why do you think it will take this time?

Because . . . .

. . . .

. . . . C844

B 1617 . . . .

page 8 90/SciM 452a

n = 215
Mean score 61%
Non-response 1%

Mark scheme

a. Key D 1

b. The planet is in between Jupiter and Uranus 1
The further from the sun the longer the trip ∴ more than 12 years but less than 84 years 1

Maximum score ③

Part a

Option	% pupils
A	5
B	5
C	2
D	81
E	6

Score for parts a and b	% pupils
3	38
2	27
1	17
0	19

Nature of response	% pupils aged 11	13	15
Pattern described at general level	1	3	5
Planet between Jupiter and Uranus in distance, ∴ between them in time	27	27	44
Planet between Jupiter and Uranus	0	9	4
Planet about half distance of Uranus from Sun, ∴ orbit time about half that for Uranus	14	16	15
Incorrect reference to other planets	14	8	3
No reference to other planets	21	8	5
Guessed—did not know	3	1	1
Irrelevant response	6	12	9

A situation in which pupils need to describe a pattern in data is when they are checking results from experiments to identify any which do not "fit" the general trend. A particular type of question based on this activity is illustrated in Example question 7. In this question the data are presented as a table of results in which one value is missing. Pupils have to determine the relationship between the variables in order to predict the missing value. In initial trials of questions of this type the majority of pupils responded by making a wild guess unless they were given some indication that a systematic method was required for a solution. The range of methods suggested in part a of the question was included for this reason but no marks were allocated to it. Overall about 60% of pupils attempted a solution by drawing a line graph. About 6% drew a bar chart and 3% attempted to find a pattern in the numbers. About 30% made no attempt to answer the second part of the question. The general levels of response to this type of question were consistent with those reported for graph drawing skills (see section 3.5.3). Between 50% and 60% of pupils set up equal interval scales. About 45% plotted the points with reasonable accuracy. Just over half of the pupils who plotted an adequate graph used it to identify an error or make the required prediction.

"Junket"

Example question 7 (4*a*)

A pudding called junket is made by adding rennin to milk to make it go solid.

Jean wanted to find out the temperature at which milk solidifies the quickest.

To do this she took eight test tubes and put into each one equal quantities of milk and rennin. She put each test tube in a beaker of water at different temperatures and timed how long it took for the milk to turn solid.

These are her results:

Test tube	Temperature in °C	Time in minutes
1	10	34
2	15	23
3	20	16
4	25	12
5	30	8
6	35	
7	40	4
8	45	4 mins.30 secs.

Jean was not able to time test tube 6 so this result is missing from the table.

a) Which one of the following would it be best for her to do in order to predict the time for milk to solidify in test tube 6?

☐ A Draw a pie chart

☐ B Work it out from the numbers

☐ C Calculate a formula

☐ D Draw a line graph

☐ E Use a histogram

b) Use the method you have chosen to predict the time for milk to solidify in test tube 6.

The time taken for milk to solidify in test tube 6 would be . . . .

n= 773
Mean score 39%
Non-response 5%

Mark scheme

a.	No marks	
b.	Attempt at drawing graph, appropriate scale	1
	Points plotted (or reasonable attempt at other method)	1
	Answer: 5 minutes± 30 seconds	1

Maximum score ③

Part a

Option	% pupils
A	57
B	7
C	3
D	8
E	25

Score	% pupils
3	24
2	16
1	14
0	47

Summary of sub-category 4α

The level of performance on questions in this sub-category varied with the type of question. In general, pupils' level of performance was higher on questions in which they had to select or make a prediction (with a mean around 50%) than for questions where a pattern had to be described (with a mean around 35%). It was clear that the level of non-response also depended on the type of question. Questions of a coded-answer type or ones requiring a very short response had non-response rates under 5%. Questions requiring a prose answer on the other hand had non-response rates between 5% and 20%.

The level of performance varied from question to question within a question type. The complexity and amount of information to be handled in the question obviously affected levels of performance. For example, questions in which pupils had to discern patterns in data involving more than two variables or which included other redundant information tended to have a lower mean score than those in which only two variables were presented.

6.3

Judging the applicability of a given generalisation: sub-category 4β

Questions in this sub-category presented pupils with data and a generalisation or hypothesis. The pupils had to assess whether or not the hypothesis was supported by the data.

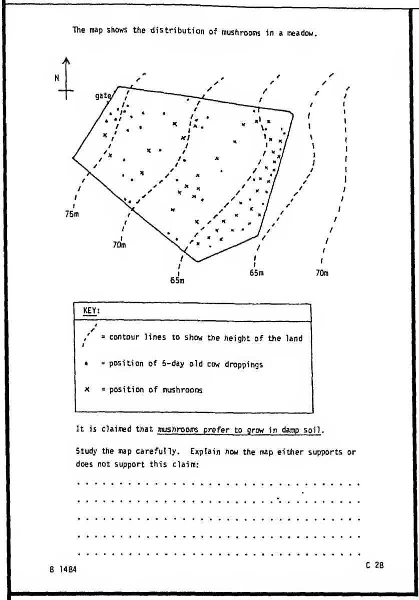
As Example question 8 shows, both relevant and irrelevant data were presented. In addition to marking pupils' responses, the different kinds of reasons given by pupils were coded. The same question was set to both 13 and 15 year old pupils. The mean score for the question was about 40% of the maximum at age 15, compared with around 30% for age 13. All pupils attempted the question and as at age 13 boys performed better than girls. (The mean score for boys was 50% and that for girls was 30%).

The types of arguments given at the two ages and the proportion of pupils giving each are shown with the question. Nearly half the pupils aged 15 attempted to relate the data given to the hypothesis. Nearly 20% simply listed observations with no attempt to relate them to the hypothesis; in some cases incorrect observations were made. Over 30% responded by making a tautological or irrelevant statement.

Four other questions of the same type were included in the survey. The mean scores were similar varying between 30% and 40%. The non-response rate on these questions ranged between 5% and 12% with about 40% of the pupils failing to score. There was no consistent difference between girls' and boys' performance on this type of question.

"Mushrooms"

Example question 8 (4β)



The map shows the distribution of mushrooms in a meadow.

N
gate
75m
70m
65m
65m
70m

KEY:

= contour lines to show the height of the land

• = position of 5-day old cow droppings

× = position of mushrooms

It is claimed that mushrooms prefer to grow in damp soil.

Study the map carefully. Explain how the map either supports or does not support this claim:

. . . .

. . . .

. . . .

. . . .

. . . .

8 1484 C 28

n = 761
Mean score 42%
Non-response 5%

Mark scheme

Supports

East end lower	1
Water will drain to it	1
More mushrooms appear to grow nearer the wall	1

If not supported

Not enough information to decide	1
The soil may be well drained and no damper at any point in the field than any other	1
There are mushrooms growing all over the meadow	1

Maximum score ③

Score	% pupils
3	27
2	18
1	9
0	45

Nature of response	% Pupils aged 13	15
Distribution of mushrooms linked to contour lines and drainage of meadow	25	29
Distribution of mushrooms linked to contour lines, no reference to drainage	10	12
More mushrooms growing on lower ground, no reference to contour lines, drainage or amount of damp	2	3
List observations only with no attempt to assess hypothesis eg. mushrooms grow near wall	15	11
List contrary or incorrect evidence (eg. more mushrooms grow near cow droppings which are source of damp)	5	3
— assume higher ground is more damp	8	4
Meaningless, irrelevant statements	20	29
Tautology	0	4

6.4 Distinguishing degrees of inference: sub-category 4γ

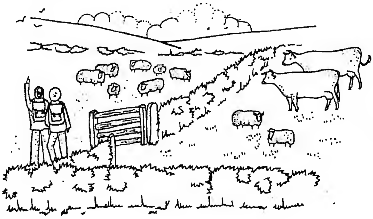
Only one type of question was used in this sub-category. Each question of this type presented an event, usually in a pictorial form, followed by statements about it made by different people. The pupils had to select the statement which made fewest additional assumptions.

Example question 9, "sheep" was used in the surveys at ages 11, 13 and 15. Results at the three ages are given for comparison. It can be seen that there is a clear increase in performance with age with about 60% of the age 15 pupils selecting the correct response. The proportion of pupils giving multiple responses decreases with age.

Three other questions of this type were included in the survey. One, "vase of flowers", showed a drawing of a vase of drooping flowers on a shelf over a radiator. Another, "dog", showed a picture of a dog with its tongue hanging from its mouth. The third, "egg shells", was set in a laboratory context. It showed a diagram of a flask containing egg shells in dilute acid and a delivery tube passing into a container of limewater.

"Sheep" Example question 9 (4α)

Look at this picture of people walking through fields in which there are animals.



Read the statements below.

Tick the one statement which you can be most sure is true just by looking at the picture.

☐ A The walkers have left the gate open.

☐ B Some sheep have strayed in with the cows.

☐ C The farmer is going to be angry.

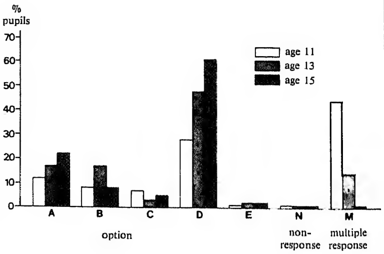
☐ D There are some sheep in both fields.

☐ E The cows will go into the other field.

A2427 C 790

n = 739
Mean score 61%
Non-response 1%

Option	% pupils
A	22
B	8
C	5
*D	61
E	2



Despite the differences in the question settings, the levels of performance on the four questions were similar (the range of mean scores was 53% – 66%).

In all these questions there were one or two distractors which were most popular. In each case these were statements which presented an obvious and plausible reason for the event. In the case of the example "sheep" the popular distractor is distractor A, "The walkers have left the gate open". It is interesting to note that although overall the level of performance on this type of question increases with age, so does the proportion of pupils who "jump to conclusions" by selecting the plausible reason.

6.5 Summary of results for data interpretation

The score distributions for the three 4α test packages (describing and using patterns in information) were similar, with means between 40% and 46% of the maximum score. Only 6 pupils failed to score at least one mark and the maximum score obtained for each test package was around 90%. The performance level of boys was higher than that for girls on this sub-category. However, this sex difference was not found in the results for the other two sub-categories, 4β and 4γ, discussed in this chapter.

Although there are considerable variations in the levels of performance on the various kinds of question concerned with data interpretation, some general patterns can be discerned. When presented with data in a range of forms about 50% of pupils can make or select a prediction based on the pattern or relationship in that data. However, only about a quarter can describe the pattern explicitly. When presented with data and a hypothesis, about 20% of pupils correctly assess the validity of the hypothesis indicating in their reasoning how it relates to the data presented. Nearly two thirds of pupils can identify a statement which makes the fewest additional assumptions about a presented event.

7 Application of science concepts

7.1 The nature of the category

It is generally accepted that one of the main aims of science education in secondary schools is to introduce pupils to a systematic framework of scientific knowledge and concepts, and to teach the use of this knowledge both in the context of school science and everyday experience. This orientation is currently reflected in much of the science teaching in secondary schools and in public examinations. The questions described in this chapter, which assess pupils' ability to use science concepts, may appear, therefore, to be the most conventional in the assessment framework.

Four sub-categories are of concern here:

$4\delta_{AB}$ Applying biological concepts

$4\delta_{CD}$ Applying physics concepts

$4\delta_{EF}$ Applying chemistry concepts

4ε Generating alternative hypotheses.

All four sub-categories require the recall and use of science concepts. The first three concern pupils' ability to make use of ideas related to three broad areas in the natural sciences. The fourth assesses their ability to use knowledge in a more divergent way by generating alternative hypotheses each of which could explain presented situations.

7.2 The concept base

A list of knowledge and concepts used in writing and classifying the questions, is given in full in Appendix 1. The list has served a useful purpose in guiding an adequate and even coverage of concepts in both the question bank and the survey tests.

During the last two years of compulsory schooling not only is there considerable variation in the content of the syllabuses in the different science subjects but (as was indicated in Chapter 2) pupils also differ according to the number and type of science subjects which they are studying. The concept list for age 15 was, therefore, developed in such a way as to reflect as far as possible the science concepts that most pupils will have been taught by the end of their third year of secondary school. The list, developed in consultation with science teachers in different parts of the country, was designed not to be prescriptive but to reflect current practice. The concepts in the list have been organised into 26 topics, which in turn have been grouped into six concept areas. For reporting purposes these areas have been paired to form three regions of scientific knowledge.

Biological knowledge:

A. Interaction of living things with their environment
B. Living things and their life processes

Physical knowledge:

C. Force and field
D. Transfer of energy

Chemical knowledge:

E. The classification and structure of matter
F. Chemical interactions

7.3 Description of the questions and tests

The types of question used to assess pupils' ability to apply science concepts were similar in style to those described in the previous chapter. Each question presented some information or a description of an event. In some questions pupils were asked either to explain the presented event or information, or to make predictions based on that information. In other questions pupils were asked to judge the appropriateness of stated conclusions.

All the questions demanded more than simple recall, and were designed to assess pupils' use of science concepts both in laboratory contexts and in situations encountered outside school.

In the next sections of this chapter the results for each of the sub-categories will be given, together with examples of questions. In order to illustrate the types of question used and the levels of pupil performance, questions have been selected from a limited number of topics in each of the three concept regions. Clearly with a total of 26 topics on the concept list it is not possible to report on each in depth on the basis of this single survey.

All questions in this category were presented in paper and pencil test booklets and were administered by teachers in the survey schools.

7.4 Applying biological concepts: sub-category $4\delta_{AB}$

The survey included 41 questions in this sub-category, distributed across three test packages. The questions had been selected at random from the question bank and were drawn from the nine biological topics as follows:

Topic	*Number of questions*
Interdependence of living things	7
The physical and chemical environment	5
Classification of living things	2
Physical and chemical principles needed to interpret life phenomena	2

The cell	2
Nutrition	3
Respiration	7
Reproduction	9
Sensitivity and movement	4

Interdependence of living things

Questions from two topics "Interdependence of living things" and "Reproduction" have been selected to illustrate the questions and the range of pupil responses obtained.

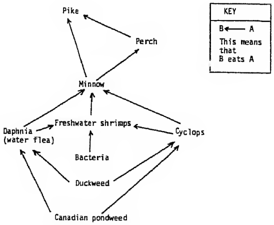
Example questions 1, 2 and 3 relate to aspects of the interdependence of living things.

Food webs can be used to represent the ways organisms depend on one another for food. In Example question 1 pupils were given a food web and asked to predict how a change in one part of the web would affect another part. 8% of pupils identified and related the two interdependent effects: the change in

"Food web"

Example question 1 (4δ)

The diagram shows the way some animals and plants living in a canal depend on each other for food.



If there is a sudden drop in the number of Cyclops, what is likely to happen to the numbers of Daphnia (the water flea)?

. . . .

. . . .

Give the reasons for your answer:

. . . .

. . . .

. . . .

. . . .

. . . .

B 1451

n =	727
Mean score	40%
Non-response	7%

Mark scheme

Increased predation *and* increased food supply	3
Increased food supply so more food for daphnia	2
Increased predation by minnow and/or shrimp so number of daphnia decreases	2
More food for daphnia	1
More daphnia eaten	1
Maximum score	③

Score	% pupils
3	8
2	45
1	6
0	41

predation and the change in the food supply. About half the pupils identified one effect only. The rest scored no marks. It is notable that, despite the key in the question, 16% of the pupils responded by inverting the direction of predation, suggesting that duckweed and pond weed eat the daphnia. The frequencies of different categories of response are given with the question. The mean score for this question was 40%: a similar pattern of response was obtained for two other questions based on food webs.

Example question 2, "animal eggs", involves the idea that predation tends to maintain the balance of populations. Only 1% of pupils mentioned the maintenance of a stable population in their response. A larger number of pupils responded by attributing purposive behaviour to the organisms being considered. The mean score was about 30% with nearly half the pupils simply selecting and restating the relevant information given in the table.

A further aspect of the interdependence of living things is the way oxygen produced by photosynthesis in green plants is used by animals and plants. Example question 3, "gas balance" requires the use of this idea. The responses were category marked to indicate the number of pupils who described the gas

"Animal eggs"

Example question 2 (4δ)

Animal	Number of offspring every season	How young are produced	Amount of parental care
Fish	More than 10,000	Lay eggs without shells	None
Amphibian	More than 1,000	Lay eggs without shells	None
Reptile	More than 100	Lay eggs with shells	Very little
Bird	5 - 10	Lay eggs with shells	A lot
Mammal	1 - 5	Give birth	A lot

This table shows the number of offspring that a pair of animals belonging to these groups produce in one breeding season. (There are, of course, exceptions.)

Why does a fish lay so many more eggs than a reptile?

Give two reasons using the information in the table above.

1 . . . .

. . . .

2 . . . .

. . . .

C 146

n = 727
Mean score 31%
Non-response 5%

Mark scheme

Eggs have no shells	1
and are more likely to be eaten	1
No parental care	1
fewer are likely to survive	1
Greater number laid to maintain population	1
Maximum score	⑤

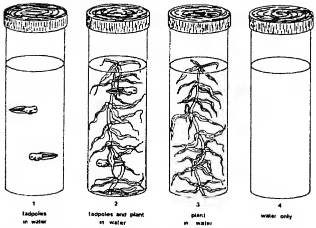
Mapped score	Raw score	% pupils
3	5, 4	10
2	3	15
1	2	32
0	1, 0	43

exchange due to photosynthesis and that due to respiration. About 12% of pupils did not respond to this question. 15% correctly predicted the gas

"Gas balance" Example question 3 (4*b*)

n = 727
Mean score 47%
Non-response 13%

Early one morning, four containers were set up as shown in the diagram. The tops were air-tight.



They were left in a well-lit place for eight hours. After this time it was suggested that the water in two of the containers would have about the same amounts of oxygen and also about the same amounts of carbon dioxide as each other.

a) Say what happens in each container.

Container One . . . .

. . . .

Container Two . . . .

. . . .

Container Three . . . .

. . . .

Container Four . . . .

. . . .

b Do you agree or disagree with the suggestion? Give your reasons.

. . . .

. . . .

Mark scheme

Container one	Oxygen removed, replaced with carbon dioxide (respiration in tadpole)*	1
Container two	Oxygen removed by tadpole and plant replaced by carbon dioxide (respiration)*	1
	Carbon dioxide removed by plant and replaced by oxygen* (photosynthesis)	1
Container three	Carbon dioxide removed, replaced with oxygen (net result of photosynthesis and respiration in plant)*	1
Container four	Balance remains the same/no change	1

If *Agree* Containers two and four contain same amount if balance assumed
OR
If *Disagree* No two containers have the same amounts if balance not assumed
1

* Give ½ mark if only one gas referred to e.g. oxygen produced.

Maximum score ⑥

Mapped score	Raw score	% pupils
3	6, 5	18
2	4, 3	32
1	2, 1	25
0	0	26

Nature of response	% pupils
Containers 2 and 3:	
Carbon dioxide removed by plant	4
Oxygen produced by plant	18
Both carbon dioxide removed and oxygen produced by plant	15
Containers 1 and 2:	
Oxygen removed by tadpoles	20
Carbon dioxide produced by tadpoles	8
Both oxygen removed and carbon dioxide produced by tadpoles	30

exchange taking place due to both respiration and photosynthesis. A further 15% mentioned only the gas exchange involved in respiration.

Reproduction

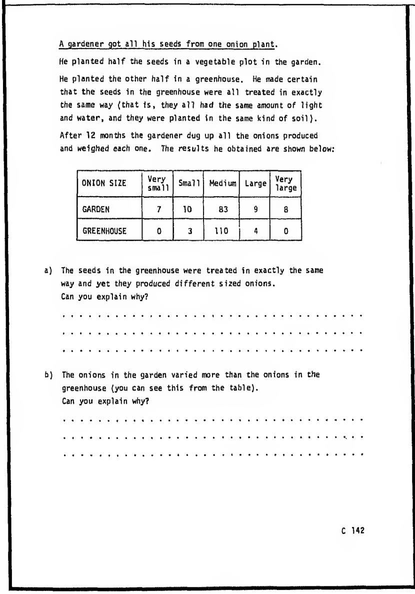
Questions in this area related to aspects of inheritance, variation, sexual and asexual reproduction in plants, and aspects of human reproduction.

The idea of variation is assessed in Example question 4, "onion plant". About 15% of pupils mentioned natural variation in their response. About a third of the pupils accounted for variation in terms of environmental conditions only, and about a quarter obtained no score.

Example question 5, "spider plant" relates to the differences in variation produced by sexual as opposed to asexual reproduction. Just under a third of the pupils selected the correct response. In Example question 6 pupils had to select from a list those instances were sexual reproduction had taken place. It appears from the responses obtained that more pupils associate sexual reproduction with animals than with plants.

"Onion plant"

Example question 4 (*4δ*)



A gardener got all his seeds from one onion plant.

He planted half the seeds in a vegetable plot in the garden.

He planted the other half in a greenhouse. He made certain that the seeds in the greenhouse were all treated in exactly the same way (that is, they all had the same amount of light and water, and they were planted in the same kind of soil).

After 12 months the gardener dug up all the onions produced and weighed each one. The results he obtained are shown below:

ONION SIZE	Very small	Small	Medium	Large	Very large
GARDEN	7	10	83	9	8
GREENHOUSE	0	3	110	4	0

a) The seeds in the greenhouse were treated in exactly the same way and yet they produced different sized onions.
Can you explain why?

. . . .

. . . .

. . . .

b) The onions in the garden varied more than the onions in the greenhouse (you can see this from the table).
Can you explain why?

. . . .

. . . .

. . . .

C 142

n = 769
Mean score 30%
Non-response 7%

Mark scheme

a.	natural variation	1
	OR	
	product of sexual reproduction/sex cells from different parents	2
b.	Environment varies more	1
	OR	
	mention of any one specific environmental factor	1

Maximum score ③

Score	% pupils
3	5
2	10
1	32
0	54

"Spider plant" Example question 5 (4δ)

A spider plant is able to reproduce in two ways:

Group 1 - by flowers and seeds

Group 2 - by producing runners (giving plantlets which root and separate from the parents)

Young plants formed by both methods were raised under identical conditions.
After 100 days each plant was weighed.
The results are shown below:

Group 1: From flowers & seeds	Mean weight 36g	Range of weights 16-55g
Group 2: From runners	Mean weight 41g	Range of weights 38-43g

Which of the following statements most fully explains the greater range of weights of plants in Group 1 than in Group 2? (Tick one box.)

☐ A Group 1 plants were raised under a wider range of environments than plants in Group 2.

☐ B All group 2 plants were heavier than group 1 plants after 100 days.

☐ C Group 2 plants contained some 'heavy' mutations, whereas group 1 plants did not.

☐ D Group 1 plants were formed by sexual reproduction, whereas group 2 plants were reproduced asexually.

☐ E Runners produce heavy and light plants in equal numbers.

C 27h

n = 727
Mean score 29%
Non-response 3%

Option	% pupils
A	15
B	19
C	18
* D	29
E	15
N	3

"Sexual reproduction"

Example question 6 (4δ)

Put a tick against each of the following in which you think sexual reproduction has taken place.

Put a cross against those in which it has not taken place.

(Make sure you put a tick or a cross in every box.)

☐ A A man planted a hundred daffodil bulbs in his garden. Each spring he cut all the flowers while they were still in bud and sold them. After a few years, he dug over his garden and found that he now had over three hundred daffodil bulbs.

☐ B A doctor took an egg cell from a woman. He fertilised the egg in a test tube using sperm from the woman's husband. He put the fertilised egg in the woman's womb where it developed. Nine months later she had a baby.

☐ C A farmer placed some day-old female chicks in a cage on their own. After a few weeks, he put each chick in its own cage and fed them till they were fully grown. They all began to lay eggs.

☐ D A sycamore tree grew in a park. Every spring, the gardener found a large number of sycamore seedlings growing in the grass around the tree.

C 772
80ScF413

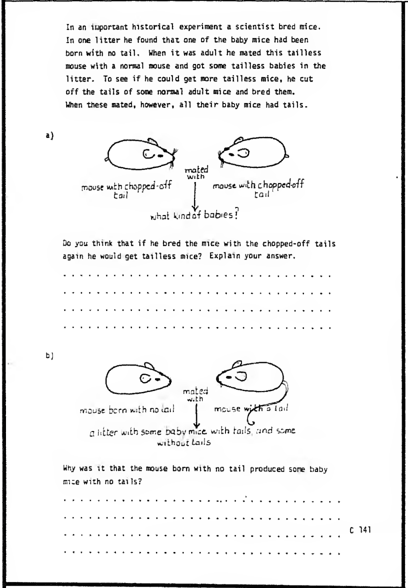
n = 707
Mean score 54%
Non-response 0%

Option		% correct
A	✗	61
B	✓	81
C	✗	77
D	✓	40

Mapped score	Raw score	% pupils
3	4	17
2	3	37
1	2	36
0	1, 0	10

"Mice"

Example question 7 (4δ)



In an important historical experiment a scientist bred mice. In one litter he found that one of the baby mice had been born with no tail. When it was adult he mated this tailless mouse with a normal mouse and got some tailless babies in the litter. To see if he could get more tailless mice, he cut off the tails of some normal adult mice and bred them. When these mated, however, all their baby mice had tails.

a)

mouse with chopped-off tail

mated with

mouse with chopped-off tail

what kind of babies?

Do you think that if he bred the mice with the chopped-off tails again he would get tailless mice? Explain your answer.

b)

mouse born with no tail

mated with

mouse with a tail

a litter with some baby mice with tails and some without tails

Why was it that the mouse born with no tail produced some baby mice with no tails?

C 141

n= 769
Mean score 74%
Non-response 1%

Mark scheme

a.	No Chopped of tail not inherited	2
b.	Mouse born with no tail passed on this inherited characteristic to some of its babies	1

Maximum score 4

Score	% pupils
3	55
2	21
1	16
0	9

Nature of response	% pupils
Part a	
Correct prediction, explanation makes specific reference to alleles (genes, chromosomes)	26
Correct prediction, explanation at a descriptive level eg. chopping off a tail is not hereditary	41
Correct prediction, incorrect or no reason given	18
Incorrect prediction	10
No response	5
Part b	
Correct prediction, explanation makes specific reference to alleles (genes, chromosomes)	32
Correct prediction, explanation at a descriptive level eg. it is passed on because it was born with no tail	53
Incorrect	16
No response	9

"Menstrual cycle"

Example question 8 (4δ)

A girl with regular menstrual cycles started to menstruate (began her "period") on the 1st of March.
She stopped menstruating on the 6th of March.
Her next menstruation began on 29th March.

On which of the following dates do you think it is most likely she released an egg from one of her ovaries?
(Tick one box.)

☐ A 1 - 6 March
☐ B 7 - 11 March
☐ C 12 - 16 March
☐ D 17 - 22 March
☐ E 23 - 28 March

n = 773
Mean score 23%
Non-response 2%

Option	% pupils
A	28
B	13
* C	23
D	11
E	22

A classical experiment on inheritance was used as a basis for Example question 7, which concerned the distinction between inherited and non-inherited characteristics. Just over a quarter of the pupils made the distinction giving an explanation in terms of hereditary material mentioning either genes or chromosomes. Nearly half of the pupils made a correct prediction and gave a more general explanation. About 10% of the pupils who responded failed to distinguish between inherited and non-inherited characteristics predicting that the mouse with the chopped-off tail will pass this characteristic on.

Example question 8 assesses pupils' knowledge of the human menstrual cycle. Fewer than a quarter of the pupils selected the correct answer to this question. About half the pupils predicted that ovulation occurs around the period of menstruation. (It is interesting to note that there were no differences between girls' and boys' results on this question.)

Summary of results for the application of biological concepts

The score distributions for each of the test packages concerning the application of biological concepts were similar with mean scores of about 40%. Eight pupils scored no marks and the maximum score obtained in each sub-test was about 90%. There was no significant difference between the scores for boys and for girls.

There were differences in levels of performance on questions of different types. These differences were consistent across the three concept regions and will be commented on later (see section 7.8).

There was little overall difference between the levels of performance on the different biological topics. The mean scores on individual topics ranged from 39% for questions on "reproduction" to 45% for questions on the "physical and chemical environment". In general, the highest level of performance was obtained for questions on the "interdependence of living things" and "the physical and chemical environment". The performance level of questions concerned with particular life processes such as respiration or reproduction was lower.

Where enough questions were included in the survey, it is possible to comment on general levels of performance on specific biological ideas.

On the topic "Interdependence of living things" over half the pupils were able to make simple predictions based on information in a food web concerning the way organisms depend on one another for food. Fewer than 20% of pupils, however, made predictions of a more complex kind in which more than one effect was considered.

Several questions related to "Sensitivity and movement". About 30% of pupils appreciated that simple animals move in response to external stimuli.

About 30% of pupils appreciated that respiration takes place in plants as well as in animals. Over half the pupils knew that oxygen is taken in by animals and used in the process of respiration with the release of carbon dioxide. The function of the blood circulation system in transporting oxygen and carbon dioxide was less well understood.

Several of the questions relating to reproduction have already been discussed. In general over half the pupils recognised that natural variation exists among members of the same species. However, fewer than a third of the pupils appreciated the idea that sexual reproduction produces greater variation than asexual reproduction. Some pupils associated sexual reproduction only with animals and not with plants.

The distinction between inherited and acquired characteristics was understood by over three quarters of the pupils.

7.5 Applying physics concepts: sub-category $4\delta_{CD}$

The survey included 45 questions in this sub-category, spread across three test packages. The questions were distributed across the seven physics topics as follows:

Topic	*Number of questions*
Movement and deformation	5
Properties of matter	10
Forces at a distance	5
The Earth in space	5
Work and energy	11
Current electricity	4
''Waves''	5

Questions relating to the topics ''Properties of matter'' and ''Work and energy'' have been selected to illustrate pupil performance within this concept region.

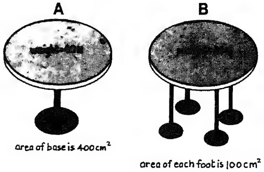
Properties of matter

Example question 9, ''tables'', required the idea of pressure for its solution. Pupils had to appreciate that both the weight of the tables and the area in contact with the carpet needs to be considered. No marks were given for the first part of the question since there is a 50% chance that pupils could obtain the correct answer by guessing: the marks were given for the reasoning. The mean score for the question was about 30%. 5% of the pupils correctly calculated the magnitude of the pressure on the carpet for each table and compared the results. A further 14% used proportional reasoning to work out a correct result. The majority of pupils considered only one variable in making their prediction, with 38% arguing in terms of the number of feet on each table. All pupils attempted the question with 2% making a prediction but giving no reason.

"Tables"

Example question 9 (4δ)

n = 750
Mean score 31%
Non-response 0%



You may have noticed that some furniture flattens the carpet.

a) Which of these tables will make the deepest impression on the same carpet?

. . . .

b) Give the reasons for your answer.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

B 1413

Mark scheme

a.	B	0
b.	twice weight/heavier	1
	same area on carpet	1
	pressure as force/area	1
	therefore B	1

Maximum score

4

Mapped score	Raw score	% pupils
3	4, 3	13
2	2	8
1	1	36
0	0	43

Nature of response	% pupils
Correct response involving force/weight and area	
Magnitude of pressure calculated for each table	5
Relative magnitude of pressures deduced correctly by proportional reasoning	14
Incorrect response relating force/weight and area	
In A: weight of 40N per foot In B: weight of 20N per foot	11
Response involving force/weight OR area	
In B area of foot snaller, therefore greater pressure	6
Weight of B greater so greater pressure	11
Weight of B on four feet, weight of A on one foot, so greater pressure on A	38
Other	13
No reason given	2

Example question 10, "car tyre", also involved the idea of pressure. In this case pupils had to select ideas related to the kinetic theory of gases and use them to explain why the pressure in a car tyre increases during a journey. The mean score for this question was again about 30%. The different types of responses given are tabulated with the question. The question was also used in the age 13 survey and the data for the two ages are presented. About 30% of age 15 pupils, (compared with 16% of age 13 pupils) give an explanation referring to particles. Less than 10% of the age 15 pupils specifically related the increased pressure to increased bombardment on the tyre in their explanation. A further 16% made no reference to particles and 29% gave an irrelevent, incomprehensible or tautological response. 4% of pupils did not attempt the question.

The third example in this section, "watering can", has been used in the surveys at all three ages. Pupils had to make predictions based on the idea that water tends to flow until the surface reaches a common level.

The mean score for the question was 42% with 29% of pupils not responding to part a. and 1% of pupils not responding to part b. The types of responses given to each part of the question are tabulated. The pattern in the results at age 15 is consistent with that at the other two ages. A notable feature is the number of pupils, nearly 50%, who did not represent the water level by a horizontal line when the vessel is tilted. In keeping with results at age 11 and 13, there was a clear difference between the level of performance of girls and boys on this question. The mean score for boys was 54% as compared with 30% for girls.

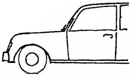
"Car tyre" Example question 10 (46)

After many experiments, scientists now think that:

- all things are made of small particles
- these particles move in all directions
- they move faster the higher the temperature
- they exert forces on each other
- they are too small to see through a microscope

Use any of these ideas to help answer the following question:

Why does the pressure in car tyres increase during a journey?



. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

C 112

n = 750
Mean score 31%
Non-response 4%

Mark scheme

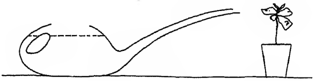
Tyres heat up	1
Air particles move faster	1
Pressure due to bombardment on wall increases	1
Maximum score	③

Score	% pupils
3	9
2	21
1	23
0	46

Nature of response	% pupils aged 13	15
Particle explanation		
Contents of tyre get warmer, air particles move faster, increasing bombardment on wall	4	9
Contents of tyre get warmer, air particles move faster, increasing pressure	10	18
Particles move faster and pressure rises	2	4
Incorrect particle explanation		
Particles produce a force on each other/move around a lot/expand	8	19
Descriptive explanations		
Heat causes air to expand so increasing pressure on tyres		
— the hotter the car the greater the pressure	2	8
— the pressure increases due to load	5	8
Other, irrelevant, incomprehensible	20	19
No attempt to explain	33	10
No response	3	4

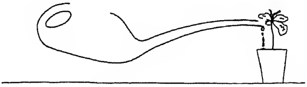
"Watering can"

Example question 11 (4δ)



The dotted line shows where the surface of the water is in this watering can.

a) Draw a line to show where the surface is in the spout.



b) The watering can is tipped so that the water just begins to drip through the spout.

Draw a line to show where the water surface is now.

B 1588 C 157

n= 761
Mean score 42%
Non-response 0%

Mark scheme

a.	horizontal line at correct level	1
b.	horizontal line at correct level in can	1
	and in spout	1

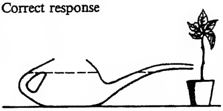
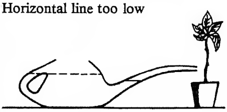
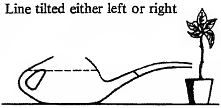
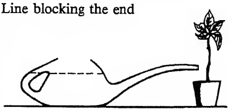
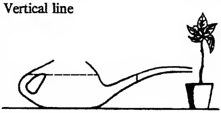
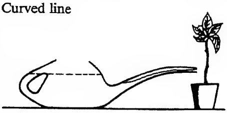
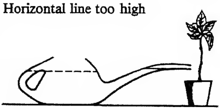
Maximum score ③

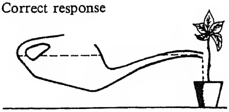
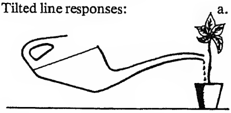
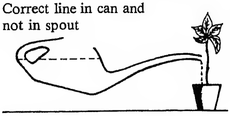
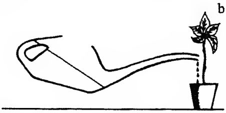
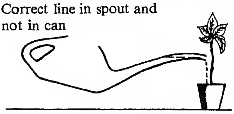
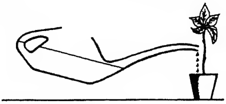
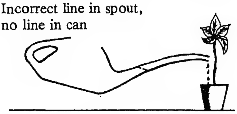
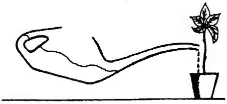
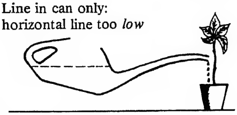
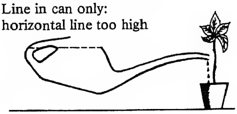
Score	% pupils
3	22
2	16
1	25
0	36

The question was included in the age 11, age 13 and age 15 surveys.

The responses to both parts of the question were categorised. The results of this are given on the next page.

"Watering can"

Part a. Nature of response	% pupils responding aged			Nature of response	% pupils responding aged		
	11	13	15		11	13	15
Correct response 	56	70	75	Horizontal line too low 	3	2	1
Line tilted either left or right 	22	14	0	Line blocking the end 	2	2	0
Vertical line 	9	0	10	Curved line 	2	0	0
Horizontal line too high 	6	9	13	Responses not included in categories or multiple lines drawn	0	3	1

Part b. Nature of response	% pupils aged			Nature of response	% pupils aged		
	11	13	15		11	13	15
Correct response 	11	21	27	Tilted line responses: a. 	16	18	16
Correct line in can and not in spout 	16	7	10	b. 	13	8	8
Correct line in spout and not in can 	6	5	5	c.  Parallel to the base	14	19	15
Incorrect line in spout, no line in can 	11	2	6		2	3	3
Line in can only: horizontal line too *low* 	3	3	4	Responses not included in the categories or multiple lines drawn	2	11	3
Line in can only: horizontal line too high 	1	2	1	No response	3	1	1

Work and energy

Four questions are discussed in this section. The first, Example question 12, asked pupils to compare the cost of heating a kettle of water to boiling point and a bath of water to just above body temperature. The question was used in the age 13 and age 15 surveys and the results for the two ages are tabulated with the question. No marks are given for the first part of the question, the prediction. Only the reason is marked. In order to compare the amount of energy expended, and hence the costs, in each case both the volume of water and the temperature rise need to be considered. This was done by only 7% of

SOUTHAMPTON LIBRARY UNIVERSITY

the pupils at age 15. 3% mentioned both factors but did not take both into account. 56% of the pupils based their prediction on only one factor, 41% only considered the greater volume of water in the bath and 8% only considered the greater temperature rise in the kettle. The mean score for the question was 27% with 2% of pupils not responding.

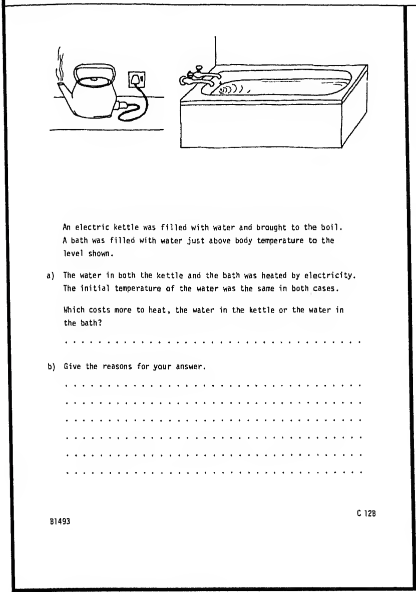
The next question, Example question 13, asked pupils to select instances in which energy is being transferred and when it is not. The mean score for the question was 69% with 1% of pupils not responding. Nearly half the pupils responded correctly to all parts of the question.

The next two questions relate to the idea of gravitational potential energy. In Example question 14 pupils need to recognise that the kinetic energy which a falling object can gain depends on the product of its mass and the height from which it is dropped. All pupils attempted the question. 32% of pupils gave the correct response to both parts of the question. Over a third of the pupils based their prediction on the height of the drop only.

The next question, Example question 15, requires pupils to recognise that the amount of energy required to raise an object from one level to a higher level does not depend on the route taken.

The mean score for the question was 11%, 2% of pupils did not attempt the question, 25% of pupils selected the correct answer to part a. but 9% of these gave a reason which either contradicted this choice or gave no reason. Over half the pupils gave responses which indicated a confusion between the notions of force applied and of energy transferred.

"Kettle and bath" Example question 12 (4*b*)



An electric kettle was filled with water and brought to the boil. A bath was filled with water just above body temperature to the level shown.

a) The water in both the kettle and the bath was heated by electricity. The initial temperature of the water was the same in both cases.

Which costs more to heat, the water in the kettle or the water in the bath?

. . . .

b) Give the reasons for your answer.

. . . .

. . . .

. . . .

. . . .

. . . .

. . . .

B1493 C 128

n = 718
Mean score 27%
Non-response 2%

Mark scheme

a.	bath	0
b.	bath contains more water	1
	kettle water raised through greater temperature	1
	ratio masses > temperature rises	1
	therefore bath costs more	1

Maximum score 4

Mapped score	Raw score	% pupils
3	4, 3	8
2	2	3
1	1	52
0	0	37

Nature of response	% pupils aged 13	15
Response involving volume of water and temperature difference		
Relative amount of water in bath and kettle related to relative temperature rise. Bath takes more energy	2	8
Responses involving volume of water OR temperatures		
Bath — more water in bath, temperatures mentioned but not taken in account	5	2
Bath — more water in bath, no mention of temperature	27	36
Bath — takes more time to heat the water therefore more electricity needed	27	7
Bath — more water in, greater heat loss	2	3
Kettle — water raised to higher temperature, relative volumes mentioned but not taken into account	3	1
Kettle — water raised to higher temperature, no mention of volumes of water	13	7
Other		11
Irrelevant or meaningless	20	23
No response	0	2

"Energy transfer"

Example question 13 (4δ)

n = 768
Mean score 69%
Non-response 1%

Here is a list of jobs.

Some require the transfer of energy, some do not.

Put a tick by those where energy is transferred.
Put a cross by those where it is not.
(Put a tick or a cross in every box.)

☐ A A person pushes a drawing pin into a board

☐ B A shelf supports a row of books

☐ C A person lifts a car using a car-jack

☐ D A crane lifts a load

☐ E A crane holds a load in position

227

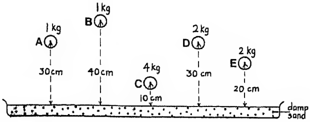
Option		% pupils correct
A	✓	85
B	×	86
C	✓	87
D	✓	70
E	×	74

Maximum score ⑤

Mapped score	Raw score	% pupils
3	5	42
2	4	30
1	3	19
0	2, 1, 0	10

"Drop sand"

Example question 14 (4δ)



Five metal balls of equal size but different masses are dropped on to a tray of damp sand from the heights shown.

a) Which will make the deepest dent?

☐ A

☐ B

☐ C

☐ D

☐ E

b) Which will make the shallowest dent?

☐ A

☐ B

☐ C

☐ D

☐ E

C 215

n = 768
Mean score 32%
Non-response 0%
Multiple response 2%

a.

Option	% pupils
A	2
B	34
C	15
* D	47
E	0

b.

Option	% pupils
* A	40
B	8
C	43
D	1
E	5

Mapped score	Raw score	% pupils
1	2	32
0	1, 0	68

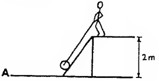
"Plank" Example question 15 (4δ)

The pictures show a man raising a heavy roller from one level to another using a rope to pull it up different planks.

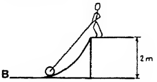
The statements below are about the amount of energy he uses in each case.

a) Put a tick in the box beside one statement you agree with.

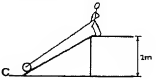
☐ A The energy he uses to lift the roller on plank A is least.



☐ B The energy he uses to lift the roller on plank B is least.



☐ C The energy he uses to lift the roller on plank C is least.



☐ D The energy he uses to lift the roller is the same whichever plank is used.

b) Give the reason for your choice of answer.

. . . .

. . . .

C 796

n= 718
Mean score 11%
Non-response 5%

Mark scheme

a.	Key D	1
b.	Lifts same weight	1
	through same vertical distance	1
	OR	
	Compensation beteen length of plank and gradient	1
		③

Score	% pupils
3	5
2	10
1	7
0	78

Nature of response	% pupils
Plank A	9
Plank B — not straight — not as steep	5
Plank C	
— smaller slope, easier/less energy to pull	22
— smaller slope	20
— other	11
Same for all planks	
— both height and weight lifted same in all cases	3
— compensation between effort used and time taken for different slopes	2
— lifts through same height OR lifts same weight in each case	9
— other	7
No reason given	11

Summary of results for the application of physics concepts

Each of the three test packages concerned with the application of physics concepts had similar score distributions, with mean scores around 32%. Eight pupils failed to score at all and the maximum score was over 90%.

There was a marked difference in the mean scores for boys and girls on this sub-category. The mean score for boys was 40% and that for girls was 30%.

There were relatively few questions relating to any particular topic and, therefore, any comparisons of performance levels between topics must be seen as very tentative indeed.

In the topic "Movement and deformation" over half of the pupils appeared to be able to interpret a speed-time graph. Fewer than 20% of pupils correctly applied dynamical concepts relating to force and motion.

About a quarter of the pupils demonstrated an understanding of air pressure and could explain gas pressure in terms of simple particle theory.

There was a higher level of performance on questions relating to electrostatics and magnetism with about 40% of pupils demonstrating a knowledge of the simple effects. However, fewer than 10% of pupils demonstrated an understanding of the simple electromagnetic effect.

On the topic "Earth in space" about two-thirds of the pupils could make predictions of relative positions and movements of objects in the solar system. However, when a quantitative element was introduced into the questions the level of performance fell.

Performance levels on questions relating to "Work and energy" were very variable. In general, questions in which pupils had to select or make simple predictions had performance levels of 50% or above. Where calculations or explanations in terms of conservation of energy or energy transfer were required the performance level was lower.

About 50% of pupils distinguished between electrical insulators and conductors. About 40% appreciated that electrical current is the same at all places in a simple circuit. Between a quarter and a third of pupils tested could apply this idea in questions involving branching circuits.

Over 50% of pupils tested made simple predictions concerning the size of shadows based on the rectilinear propagation of light. About one third of the pupils applied the law of reflection of light to an everyday problem, and about 10% indicated the general direction light would travel as it passed from air to water.

7.6 Applying chemistry concepts: sub-category $4\delta_{EF}$

41 questions were included from this sub-category distributed across three test packages. The number of questions selected from each of the ten topics in this region was as follows:

Topics	*Number of questions*
States of matter	5
Pure substance	4
Metals and non-metals	3

Acids and bases	4
Periodic table	2
Atomic model	2
Solutions	5
Reactivity	2
Properties of a chemical reaction	7
Some chemical reactions	7

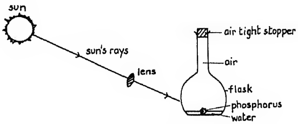
Questions related to the topics "Properties of a chemical reaction" and "Some chemical reactions" have been selected to illustrate pupil performance within the concept region. Additional questions relating to various aspects of the "Classification of matter" have also been included.

Properties of chemical reactions

General properties of chemical reactions, the identification of reversible as opposed to irreversible changes and factors affecting rates of reaction were included in this topic. Example question 16 concerned the conservation of mass during a chemical reaction in which phosphorus was burnt in a sealed flask. As the results indicate, about a third of pupils predicted correctly that the mass of the contents of the flask would be the same before and after the reaction. The most common response, given by 45% of pupils, was to predict a decrease in the mass of the contents of the flask. 13% predicted an increase in mass. The various reasons given by pupils at ages 13 and 15 are tabulated with the question. The pattern of the responses was similar at the two ages. The analysis of the incorrect responses indicates some common errors. For example, nearly 10% of pupils suggested that a change of mass occurs on dissolving. There is also considerable confusion between the ideas of mass and density; for example, volume changes which occur with changes of state are also associated with mass changes. The mean score for boys was higher than that for girls; a result found also at age 13.

"Phosphorus"

Example question 16 (4δ)



A piece of phosphorus was held in a flask as shown in the diagram. The mass of the flask and contents equalled 205 g. The sun's rays were focussed on the phosphorus, which then caught fire. The white smoke produced slowly dissolved in the water.
After cooling, the flask and its contents were weighed again.

a) Would you expect the weight to be:

☐ A More than 205 g
☐ B 205 g
☐ C Less than 205 g
☐ D Not enough information to answer

Tick in the box next to the answer you choose.

b) Give the reason for your answer:

. . . .

. . . .

. . . .

C 221

n =	777
Mean score	25%
Non-response	2%

Mark scheme

a.	Key B	1
b.	Mass/matter cannot be created or destroyed	1
	Mass remains constant	1
	OR	
	Nothing enters or leaves	1
	Mass remains constant	1

Maximum score ③

Part a

Option	% pupils
A	13
*B	33
C	45
D	8

Score	% pupils
3	14
2	11
1	9
0	65

Nature of response	% pupils age 13	% pupils age 15
Option A selected: mass increases		
Size of phosphorus increases on burning hence mass increases	5	2
Gas produced heavier than air or gas dissolves making water/flask heavier	7	5
Irrelevant/no response	4	6
Option B selected: mass remains same		
Nothing enters nor leaves flask so mass remains same	6	22
Phosphorus would lose weight but the product dissolves in water	2	4
Smoke produced too light to weigh anything	16	1
Irrelevant	6	6
Option C selected: mass decreases		
Burnt phosphorus or smoke weighs less/less dense than solid	20	20
Phosphorus dissolves so only water is left	9	8
Water lost/evaporated	9	2
Phosphorus lighter because melted or took in oxygen	4	5
Irrelevant/no response	5	10
Option D selected: not enough information	7	8

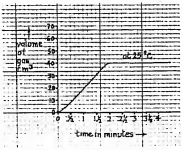
Example question 17 is concerned with the change in rate of reaction with temperature. Nearly half the pupils selected the incorrect option predicting an increased volume of product with an increase in temperature. About a third of pupils selected the correct option in which the rate of the reaction changes but not the volume of the product.

In a parallel question pupils were given the same information but asked to select which graph they would expect if the initial volume of the solution was halved. In this case nearly two thirds selected the correct response. Other questions also indicate that many pupils confused the rate at which a reaction occurs and the mass or volume of the reaction products.

"Reaction rate"

Example question 17 (4δ)

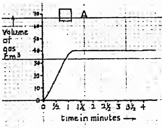
When an excess of solid was added to a solution at 25°C a gas was given off. The volume of gas measured against time was plotted below:

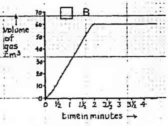


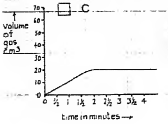
The same experiment was repeated with the solution at 35°C.

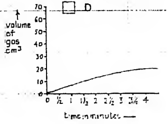
Which one of the following graphs would you expect as a result?

(Tick one box)









n = 777
Mean score 31%
Non-response 7%

Option	% pupils
*A	31
B	49
C	9
D	4

In a parallel question, pupils had to select which graph they would expect if the initial volume of solution was halved.

The results were as follows:

Option	% pupils
A	13
B	9
*C	63
D	9

Some chemical reactions

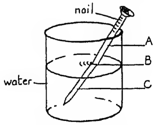
Several of the questions on this topic related to the reaction of metals with oxygen.

Example question 18 concerns the rusting of an iron nail. About half the pupils selected the correct response and about a third indicated that both air and water were necessary for rusting to take place. In another question pupils were shown pictures of an iron bolt left in three environments. They had to predict whether the bolt would rust, in each case giving their reasons. In this question about half the pupils identified air and water as necessary for rusting to occur.

"Nail"

Example question 18 (4δ)

A bright, shiny iron nail was left for a few days in a jam jar half-full of water, as shown in the diagram below.



a) Where would you expect the nail to rust most?

☐ A At A

☐ B At B

☐ C At C

☐ D It would rust evenly all over

b) Give the reasons for your choice of answer.

. . . .

. . . .

. . . .

. . . .

C 268

n =	707
Mean score	34%
Non-response	1%
Multiple response	2%

Mark scheme

a.	Key B	1
b.	Air and water both necessary	1
	Maximum where two meet at surface	1

Maximum score ③

Part a

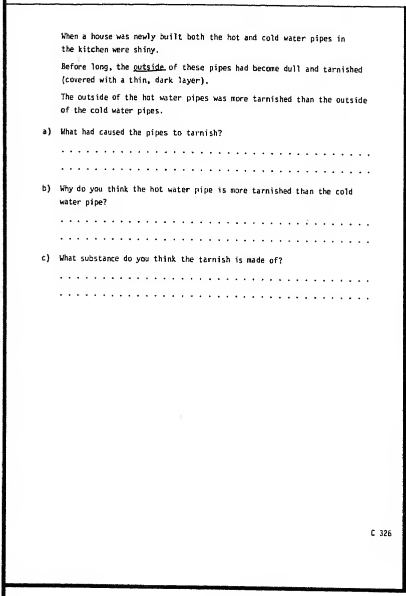
Option	% pupils
A	6
*B	53
C	33
D	4

Score	% pupils
3	8
2	30
1	20
0	42

A lower level of response was obtained on a question concerning the reaction of copper and oxygen, Example question 19. In this question only about a fifth of the pupils appreciated that the black coating on the copper pipe was due to a reaction between copper and oxygen from the air.

"Copper pipes"

Example question 19 (4δ)



When a house was newly built both the hot and cold water pipes in the kitchen were shiny.

Before long, the outside of these pipes had become dull and tarnished (covered with a thin, dark layer).

The outside of the hot water pipes was more tarnished than the outside of the cold water pipes.

a) What had caused the pipes to tarnish?

. . . .

. . . .

b) Why do you think the hot water pipe is more tarnished than the cold water pipe?

. . . .

. . . .

c) What substance do you think the tarnish is made of?

. . . .

. . . .

C 326

n = 707
Mean score 19%
Non-response 11%

Mark scheme

a.	reacts with oxygen from air	1
b.	heat increases rate of reaction	1
c.	copper oxide	1

Maximum score ③

Score	% pupils
3	10
2	9
1	9
0	72

Nature of response	% pupils
Part a	
Reacts with air/oxygen	16
Response includes mention of air around pipes but no mention of reaction	14
Other responses (including dirt, steam, condensation)	29
Heat or cold	20
Part b	
Temperature affects speed of reaction	16
Some mention of temperature difference but not related to reaction	42
Part c	
Copper oxide or oxide of metal	22
Rust	8
Other substances (dust, condensation)	38

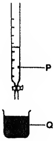
Classification and structure of matter

In this section questions from a range of topics including "Acids and bases", "Atomic model" and "States of matter" are discussed.

Example question 20 required two points to be taken into account: that the liquid Q was neutral initially and that it became more acidic when liquid P was added. Between 40% and 50% of pupils selected responses which reflected appreciation of each of these points separately. 30% selected the alternative which included both of them.

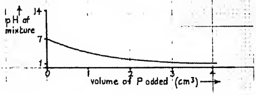
"pH graph" Example question 20 (46)

A pupil has this apparatus:



- she measures the pH of Q
- she adds P to Q 1 cm³ at a time
- after each addition she measures the pH of Q

This is a graph of her results:



What does the graph show the pupil was doing?

Put a tick in the box by the one you choose.

☐ A Adding acid to alkali
☐ B Adding water to alkali
☐ C Adding acid to water
☐ D Adding alkali to water
☐ E Adding alkali to acid

C 16P

n =	772
Mean score	30%
Non-response	3%
Multiple response	0%

Option	% pupils
A	19
B	16
*C	30
D	12
E	20

In Example question 21, pupils had to predict the atomic species which is produced when an electron is removed from a sodium atom. About a third of the pupils selected the correct response. Over a third predicted that sodium would change to another element and nearly a third predicted the wrong charge on the sodium ion.

In another question in which pupils had to state the number of protons and electrons in atoms where the atomic number was given, about a third of the pupils tested answered correctly.

Summary of results for the application of chemistry concepts

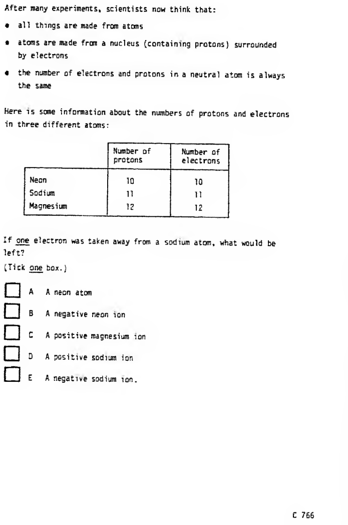
In each of the three test packages scores ranged from 0 to 94% with just 11 pupils failing to score. The sub-test score distributions were similar with a mean about one third along the mark scale.

There was no significant difference between the mean scores for boys and girls on this sub-category.

Responses to individual questions indicated that between 30% and 50% of pupils applied the idea that the temperature of substance remains constant when changes of state occur.

"Sodium"

Example question 21 (4*b*)



After many experiments, scientists now think that:

- all things are made from atoms
- atoms are made from a nucleus (containing protons) surrounded by electrons
- the number of electrons and protons in a neutral atom is always the same

Here is some information about the numbers of protons and electrons in three different atoms:

	Number of protons	Number of electrons
Neon	10	10
Sodium	11	11
Magnesium	12	12

If one electron was taken away from a sodium atom, what would be left?

(Tick one box.)

☐ A A neon atom

☐ B A negative neon ion

☐ C A positive magnesium ion

☐ D A positive sodium ion

☐ E A negative sodium ion.

C 766

n = 772
Mean score 33%
Non-response 3%
Multiple response 1%

Option	% pupils
A	17
B	11
C	8
*D	33
E	27

About 40% of pupils appreciated the function of an indicator with about 30% correctly interpreting the pH scale.

On the topic, "Solutions", between 30% and 40% of pupils demonstrated an appreciation of the idea of concentration as a ratio. About 40% of pupils indicated an appreciation that heating tends to increase the solubility of solids in liquids.

There was considerable variation in the mean scores for questions related to the topics "Properties of a chemical reaction" and "Some chemical reactions". 70% of pupils made simple discriminations between reversible and irreversible reactions. About 30% of pupils distinguished between the rate of a chemical reaction and the volume or mass of the products. Again, about 30% demonstrated an appreciation that in a chemical reaction the total mass of the reacting substances is the same as the total mass of the products.

On questions concerned with specific chemical reactions nearly 70% of pupils knew that oxygen was needed to sustain combustion and about half identified air and water as necessary for rusting to take place.

7.7 Generating alternative hypotheses: sub-category 4ε

In questions of this kind pupils are presented with a description of an event and they are required to suggest up to three possible hypotheses which could account for the observations. The purpose of these questions is not to elicit the correct reason, the events being as far as possible novel ones for which few pupils would be expected to know the accepted explanation, but to assess the number of plausible hypotheses which pupils can generate.

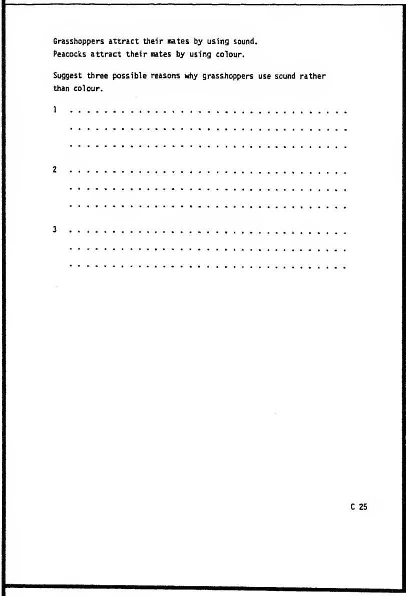
Pupils are asked to suggest up to three hypotheses. To be marked correct any hypothesis must be consistent with the data presented, and be in keeping with scientific concepts and principles.

Example questions 22 and 23 are of this type. For each of these questions, the hypotheses suggested by pupils have been categorised to give an indication of their range and variety.

Example question 22 "Grasshoppers" had the lowest mean score of all the questions of this type. In many cases pupils failed to relate their suggested hypothesis to the observations. Suggestions of this kind included "grasshoppers are not beautiful", "grasshoppers are too small", "noise travels farther". As might be expected the highest number of adequate hypotheses is given in the first attempt with about 5% of pupils restating a previous hypothesis for the second and third attempts.

Example question 23 "Ivy" was used in the surveys at ages 11, 13 and 15. The mean score for the question was high with 82% of all attempts being acceptable hypotheses. This level of response was similar to that obtained by 13 year old pupils and higher than for 11 year olds. The mean score for the question was 49% as compared with 44% and 30% at ages 13 and 11 respectively. This indicated that although the number of acceptable hypotheses was similar at ages 13 and 15, the hypotheses at age 15 tended to be related more explicitly to the observations.

"Grasshoppers" Example question 22 (4ε)



Grasshoppers attract their mates by using sound.
Peacocks attract their mates by using colour.

Suggest three possible reasons why grasshoppers use sound rather than colour.

1 ....
....
....

2 ....
....
....

3 ....
....
....

C 25

n = 769
Mean score 27%
Non-response 5%

Mark scheme

Any scientifically plausible hypothesis 1 each

Maximum score ③

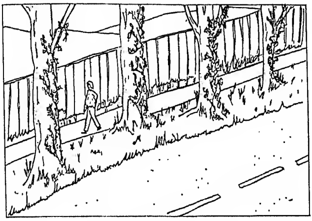
Number of acceptable hypotheses	% pupils
3	1
2	16
1	46
0	37

Nature of response	% pupils giving hypotheses to part a	b	c
Statements including reasons			
Bright colours make grasshoppers more visible to predators (dull colours camouflage)	25	15	8
In grasshoppers' environment sound can be detected at a greater distance than they can see colour	7	7	4
Grasshoppers mate only at night/can be heard when cannot be seen	0	0	1
Statements with reasons inferred eg. grasshoppers are colour blind grasshoppers have well developed hearing grasshoppers have mechanism for making characteristic sounds	10	16	16
Statements giving no reason eg. grasshoppers are too small grasshoppers are not beautiful	50	31	18
Repeat of previous statement	—	4	5
Incomprehensible response	5	13	17
No response	5	11	28

"Ivy"

Example question 23 (4ε)

Walking along this footpath, Thomas noticed that there was ivy growing on the trees, but only round three-quarters of the trunks. None of the trees had ivy growing on the side nearest to the path.



Think of two different reasons why the ivy might grow only on some sides of the trees.
Write the first at a) and the second at b).

a) I think it might be because . . . .

b) or it might be because . . . .

A 2469

C 791

n = 769
Mean score 42%
Non-response 6%

Mark scheme

For each part a and b	
Acceptable hypothesis and explanation	2
Acceptable hypothesis	1
Maximum score	**4**

Score	% pupils
4	10
3	26
2	37
1	13
0	13

Nature of response	% pupils age 11	 age 13	 age 15
Acceptable hypotheses			
Not so much light on the side shaded by the fence	8	12	12
Sunlight more on one side	14	19	36
Sun on path side and would dry up ivy	2	3	1
People/animals on path pull/knock off ivy	11	14	5
Not enough water on one side because of fence	6	10	5
Wind blows harder on path side blowing ivy back	5	4	10
Fungus or other factor on path side stops ivy	3	5	1
Vehicles on road provide things ivy needs	2	5	4
Soil is poorer on one side	2	3	2
Ivy only planted on one side	6	6	1
Has not had time to grow on other side	4	2	1
Temperature higher on road side encouraging ivy growth	0	0	4
Observations only	2	2	2
Reasons for which there is no evidence or contradictory evidence	10	5	4
Irrelevant/tautological/meaningless statements	21	7	6
Non response	0	5	6

Six questions of this kind were included in the survey two in each of the three concept regions. The distributions of scores for the questions were similar. About 10% of pupils generated three plausible hypotheses, 40% generated two and 30% generated one. About 20% of pupils failed to score, with around 4% – 8% not attempting the questions.

It is of interest that on each of the six questions of this type used in the survey the mean score for the boys was higher than that for the girls. The overall mean score for the boys was 44% and that for the girls was 39%.

7.8 General comments on the interpretation of question scores

In general the level of performance varied with the type of question. The same pattern emerged across the three concept regions (biology, physics and chemistry). Questions in which pupils selected a response had a generally higher mean score than those requiring an open-ended response. (This could be due to the lower non-response rates to these questions and to some pupils guessing the correct response.) Questions in which pupils had to assess a presented hypothesis appeared to have the lowest mean scores overall.

The overall sub-category scores indicate that the mean score for boys was higher than that for girls in the physical science concept region. Questions where boys performed at a higher level tended to be those involving the explanations of everyday phenomena or practical applications. On questions of a more abstract kind involving the interpretation of tabulated or graphical data the difference between the sexes was not so marked.

The comments made at the end of each section in this chapter about levels of understanding of particular topics are based on relatively few questions. We know that the context and form of a question can affect levels of performance. The results reported in this survey should be seen as tentative and care must be taken not to generalise from them at this stage. When more questions on a topic have been used in future surveys it may be possible to report on performances within a topic in more detail.

8 Design of investigations

8.1 Introduction

The careful planning of investigations is fundamental to much scientific work, and the general principles involved in such planning, for example the control of variables and the sequencing of procedures, are applicable in many areas of enquiry. Designing investigations is, therefore, an important aspect of scientific activity though it often seems to be neglected both in teaching and in assessment.

An experiment can only be designed if a problem or hypothesis is framed so that it is open to empirical investigation. This idea forms the basis of the first sub-category, 5α, in which pupils are asked either to assess whether statements are testable or not, or to rephrase a vague or ambiguous statement into one that could be experimentally investigated. Having isolated a problem, an experimental procedure has to be devised to solve it, and the second sub-category, 5β, is concerned with assessing this skill. This sub-category includes a variety of question types which test different aspects of experimental design from control of variables to putting a number of experimental steps into the correct sequence.

The third group of questions in this category is different from any other used in the science survey. These questions call for extended pieces of writing with pupils being asked to describe in detail what they would do to solve a particular problem or to test a given hypothesis. Since only a few questions of this type can be included in any one survey, they are reported as individual questions, and no overall sub-category score is computed. These longer questions encompass many of the elements that are tested separately in sub-category 5β, and also overlap with Category 6, the individual practical investigations. Category 6 was not tested at age 15 in the 1980 survey; in the 1981 survey the differences in performance between what pupils write when asked to design an experiment, and what they actually do practically when presented with the same problem, is being investigated.

This chapter first describes the main question types used to assess the sub-categories, with example questions shown, and then discusses performance levels on the two sub-categories. The final part of the chapter describes pupils' responses to the longer questions.

8.2 Assessing testable statements: sub-category 5α

8.2.1 Introduction

It is difficult to investigate a problem empirically until the problem itself is clearly defined. Questions in this sub-category ask pupils to assess whether statements are open to scientific investigation or not. Some statements, of the "lead is very heavy" variety, are not defined with sufficient precision to be testable; others, typified by advertising material, contain too many unspecified related variables to allow testing of the statement as it stands. There were two types of questions used in this sub-category: "identifying testable statements" and "rephrasing a general statement".

8.2.2 Identifying testable statements

In these questions, pupils had to select which of a group of 4 or 5 statements was *not* open to scientific investigation.

A question of this type is shown in Example question 1.

Identifying testable statements

Example question 1 (5α)

Below are some things people say about powdered potato and fresh potato.

All except one of these statements can be checked to find out if they are true.

Put a cross in the box by the one that cannot be checked scientifically.

☐ A Powdered potato is more useful than fresh potato.

☐ B Powdered potato has less vitamin C than fresh potato.

☐ C Powdered potato has more calories than fresh potato.

☐ D Powdered potato has more water in than fresh potato.

☐ E Powdered potato has less protein than fresh potato.

C 723

n = 755
Mean score 74%
Non-response 0%
Multiple response 1%

Option	% pupils
*A	74
B	5
C	4
D	10
E	5

15 of these questions were used altogether, 5 in each of three test packages. They are relatively novel since in most tests pupils expect to make a choice on whether an answer is a correct statement, not whether it is a testable one. The response rate was high on these questions, fewer than 1% of pupils not attempting an answer. However, a few pupils (1 – 3%) marked more than one box despite the instructions in the questions.

Mean scores ranged from 20% to 90%, with the effect of the science content in the questions being most striking. When a question was set in a science context, or where science knowledge was required to answer a question, the mean score was in the range 20% to 40%; questions which were set in everyday situations and which only needed a recognition of the vagueness of words such as "better" had mean scores which ranged from 60% to 90%. This difference in mean scores could be because the kinds of tests required in the everyday situations are more familiar to most pupils, so they are better able to judge which statement is untestable.

Some of the questions in this group deliberately had distractors which were easily testable, but which were incorrect statements. Option D in question 1 is an example of this. "Powdered potato contains more water than fresh potato" is not a true statement but is testable. Around 10% of pupils opted for these kinds of distractors.

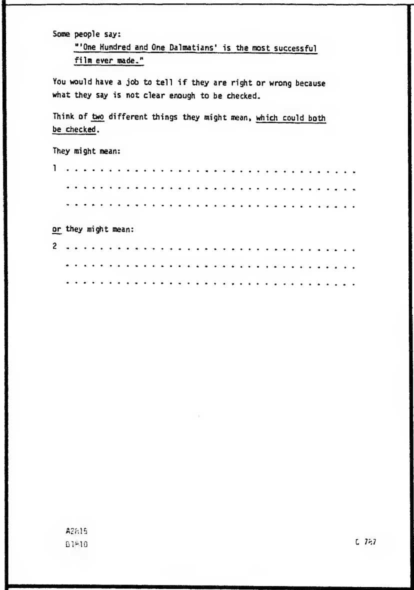
8.2.3 Rephrasing a general statement

In these questions, pupils were given a statement which was too vague or ambiguous to be testable. They were asked to rephrase it in a way that would make it testable.

The following question is an example of this type, and was also used in the 1980 survey of 13 year old pupils (DES 1982a).

Rephrasing a general statement

Example question 2 (5*α*)



Some people say:

"'One Hundred and One Dalmatians' is the most successful film ever made."

You would have a job to tell if they are right or wrong because what they say is not clear enough to be checked.

Think of two different things they might mean, which could both be checked.

They might mean:

1 ....

....

....

or they might mean:

2 ....

....

....

A2R15
B1R10

C 7R7

n = 765
Mean score 52%
Non-response 9%

Mark scheme

More people went to see 101 Dalmations than any other film	1
More cinemas hired it than any other film	1
It made a bigger profit	1
The most popular film	1
It won more awards than.....	1
It ran longer than.....	1

Any 2 statements Maximum score ②
NB. Answers must be restatements

Score	% pupils Age 13	% pupils Age 15
0	58	48
1	22	32
2	20	20
Non-Response	3	9

Mapped score	Raw score	% pupils
1	2, 1	52
0	0	48

Discussion

The mean score for this question is about half-way along the range for this question type.

There is little difference in performance between ages 13 and 15 on this question.

9 questions of this type were used, spread across the three test packages.

All the rephrasing questions were set in everyday situations, and although the skill demanded by each question is the same, the mean scores ranged from 35% to 77%. This could be due to the fact that in some questions pupils found it easier to think of the experiment they would need to do to test the statement.

Examples of acceptable responses are shown on the mark schemes. For each question, mark scheme examples were based on results from question trials, but markers were instructed to accept alternatives as long as they were testable statements.

Most of these questions were also coded to indicate if the answer the pupil gave was a simple rephrasing or a rephrasing which included the idea of controlling the test situation.

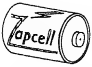
Some of the pupils failed to rephrase, but briefly described what they would do to test whether the statement was true. This type of response was also coded, and scored marks if the test would work. Example question 3 was marked in this way.

Rephrasing a general statement

Example question 3 (5α)

The shop assistant said:

"Zapcell batteries last longer."



You would not be able to check if this statement is true because it is too vague.

Think about what it could mean.

Write a statement of your own which means the same thing, but puts it in a way that could be checked.

. . . .

. . . .

. . . .

. . . .

C 742

n =	741
Mean score	67%
Non-response	6%

Mark scheme

Zapcell batteries last longer than other kinds of batteries	1
OR	
Similar statements that can be checked	1
Any 1 statement Maximum score	①

Rephrased	46%
Rephrased with controls	5%
Described workable test	15%
Described unworkable test	4%

This was one of the easiest questions of this type.

In Example question 3, where pupils were asked to rephrase the statement "Zapcell batteries last longer", those who rewrote the statement with the idea of controls used phrases such as "Zapcell batteries last longer than other batteries when tested in the same kind of torch".

A common example of a correct test from the battery question was "Put Zapcell in one torch, and another battery in another torch, and see which lasts longer".

The number of pupils who rephrased with controls varied from less than 1% up to 6%. The number who wrote about a test, rather than rephrased the statement varied from less than 1% to 32%. This may bear some relation to the simplicity of the test required.

Extensive trials were carried out with this type of question, to attempt to make the question stem as comprehensible as possible. Extended interviews were also conducted with a number of pupils. From these, it appears that seeing the vagueness in a statement, and being able to rephrase it, is a skill that is not easy for pupils to acquire. Most of those who had failed to score originally, failed to score again even after examples had been discussed with them.

Those pupils who attempted a question but failed to score made a variety of errors — from simply recopying the statement or changing the order of the words to substituting another untestable word, or writing a statement which was barely related to the problem. For example, in rephrasing the statement "The weather is worst at weekends" a fairly common type of response was "The man is at work all week and the weather doesn't bother him then." In Example question 3 "Zapcell batteries last longer" a common response was "Zapcell batteries are best".

As with other questions where pupils have to commit their own ideas to paper, the response rate on this question type was lower than for coded-answer questions: between 5% and 13% of pupils failed to make any attempt.

8.3 Assessing experimental procedures: sub-category 5β

8.3.1 Introduction

If their experiments are going to investigate problems satisfactorily, pupils need to know which variables they must keep the same, which they need to change, which are important, which are irrelevant and the sequence of procedures needed to perform the experiment. They should be able to criticise a given procedure, know what the different activities or parts of the apparatus are for, and know what their experiment will and will not test. This sub-category contains a wide variety of questions, covering all these aspects involved in assessing experimental procedures. The questions vary in the amount of taught science knowledge needed to answer them, and in the context in which they are set, some being in traditional laboratory settings, and others in everyday situations. Some questions are based on experimental procedures one would expect pupils to be familiar with because they are commonly taught in schools, others are based on novel situations, in both laboratory and everyday settings, where pupils have to apply the same general principles.

The 54 questions used to assess pupils' performances on experimental procedures were selected randomly from all the questions on this topic. Seven types were represented, but most of the questions came from five main question types, each of which looked at a different aspect of assessing procedures.

8.3.2 Selecting variables to be controlled, or experimental procedures necessary for solving a problem.

This was the largest group of questions, and also the most varied in style. Pupils were asked to select the best experimental procedure for solving a given problem. In most cases this involved working out which variables needed to be kept the same and which needed changing to allow the problem to be solved.

The following two questions are examples of this type. They also illustrate the fact that drawings can form an important part of the questions. The second was also used in the 1980 survey of 13 year olds (DES 1982a).

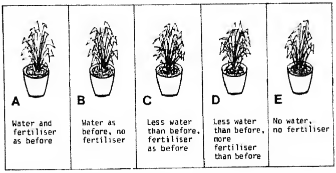
Controlling variables

Example question 4 (5β)

A lady buys five household plants of the same kind.

- She puts them side by side on a windowsill.
- Each day she gives them water and a little liquid fertiliser.

All the plants begin to wilt.
She wants to find out why they are wilting.
She treats her plants as shown below.


A
Water and fertiliser as before
B
Water as before, no fertiliser
C
Less water than before, fertiliser as before
D
Less water than before, more fertiliser than before
E
No water, no fertiliser

a) Which two does she need to compare to find out if the plants wilted because she had given them too much water?

She should compare ☐ and ☐

b) Which two does she need to compare to find out if the fertiliser was making them wilt?

She should compare ☐ and ☐

c) Which two does she need to compare to find out if the fertiliser is helping them to grow?

She should compare ☐ and ☐

C 6.24

n = 741
Mean score 42%
Non-response 2%

Mark scheme

	Response	Mark	% pupils
a.	A and C	1	38
b.	A and B	1	51
c.	C and D	1	38

Maximum score ③

Raw score	% pupils
3	17
2	22
1	32
0	29

Of those who attempted this question, 29% did not score, and 17% got all three parts correct.

All other possible combinations were present in the range 2% – 14%.

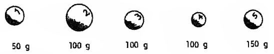
The most frequent were to get b only or a and b right. The least frequent were to get a only or a and c right.

Selecting a procedure

Example question 5 (5β)

You are asked to do an experiment to find out if the volume of a ball makes any difference to the time it takes to fall to the ground.

You are given a box containing the balls shown in the diagram.



Which one of the following would you do?
(Tick one box.)

☐ A Drop balls 1, 3 and 5 from the same height and time how long each takes to fall.

☐ B Drop ball 3 and time how long it takes to fall from different heights.

☐ C Drop balls 2, 3 and 4 from the same height and time how long each takes to fall.

☐ D Drop balls 1, 3 and 5 from different heights and time how long each takes to fall.

C 527

n = 765
Mean score 44%
Non-response 1%
Multiple response 1%

Mark scheme

Response	% pupils age 13	% pupils age 15
A	31	40
B	4	6
*C	48	44
D	10	9
M	5	1
N	1	1

At age 15, the mean score for girls was 38%, for boys 48%, but this difference is not characteristic for the question type as a whole.

In option A, the volume is held constant when it is the variable which should be changed.

While the pattern of responses is the same at both ages, the mean score at age 13 is slightly higher. There was one difference between the questions — at age 15 the word "volume" was used at age 13 the word "size" was used.

19 questions of this type were used. In each case the pupils had to respond by putting a tick or letter in a box. Mean scores ranged from 30% – 90%. As was the case in 5α the need for the knowledge of taught science concepts or specific procedures had a noticeable effect on the mean score of a question. Questions which required the application of general principles but without the need for specific science knowledge had mean scores of 40% – 90%. Questions which test standard procedures (such as chromatography or testing expired air for component gases) and those involving taught science concepts had mean scores of 30% – 50%.

The response rate on these questions was uniform and generally high; fewer than 2% of pupils made no attempt on ordinary multiple choice questions, and 2% – 4% on questions where more than one box had to be marked in some way. About 1% of the pupils scored zero through marking more than one box, and 3% – 7% marked only one box when two or more should have been marked.

There was no overall difference in the performance of boys and girls on this question type, although there was some variation on individual questions. This did not follow any obvious pattern, girls scoring as well as boys on several questions which had a physical science bias. Boys appear to be slightly better at questions involving standard laboratory procedures.

8.3.3 Stating which variables need to be controlled

In this type of question, pupils were given details of a problem, and the equipment and method selected to solve it. They were asked to suggest what should be done to create a "fair test". This type of answer therefore required a limited amount of writing.

In some questions of this type that involved use of science concepts, mark schemes had to be on two levels to take account of the relative importance of the variables mentioned. Some pupils answered from knowledge, mentioning the most important variables, while others suggested other correct, but much less critical, things that should be done to make the test fair.

This was the case in Example question 6, where examples of lower level responses can be seen in the mark scheme.

Controlling variables

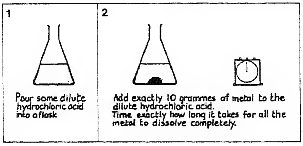
Example question 6 (5β)

n = 765
Mean score 32%
Non-response 19%

When some metals react with dilute hydrochloric acid, hydrogen gas is given off.

A pupil wants to find out if metal A reacts faster with dilute hydrochloric acid than the same mass of metal B.

This is what he plans to do for each metal:



Pour some dilute hydrochloric acid into a flask

Add exactly 10 grammes of metal to the dilute hydrochloric acid.
Time exactly how long it takes for all the metal to dissolve completely.

Write down three other things he must do to make sure the test is fair.

1 ....
....

2 ....
....

3 ....
....

C 705

Mark scheme

1. Use same volume (amount) of acid
2. Use same concentration (strength) of acid
3. Use metal in same form (same surface area)
4. Stir or shake both or neither 2
5. Control temperature of each

Or

Any other *good* response

Following are examples of lower level responses which can score *1 mark*

6. Keep all the conditions the same
7. Use clean/new acid each time
8. Use clean/washed flasks each time
9. Stand on a white background
10. Judge dissolving point accurately
11. Use same size flasks 1
12. Make sure metal is completely covered

Or

any other valid lower level answer

Any 3

Maximum score ⑥

Note: Errors of timing or weighing are not acceptable since they are controlled in question stem.

Mapped score	Raw score	% pupils
3	6	8
2	5, 4	20
1	3, 2	30
0	1, 0	42

Discussion

This question requires knowledge of the factors which affect chemical reactions, and was one of the hardest of this question type.

The non-response rate was also high.

The mean score for girls was 28%, for boys 35%, but this was not the case on other similar questions.

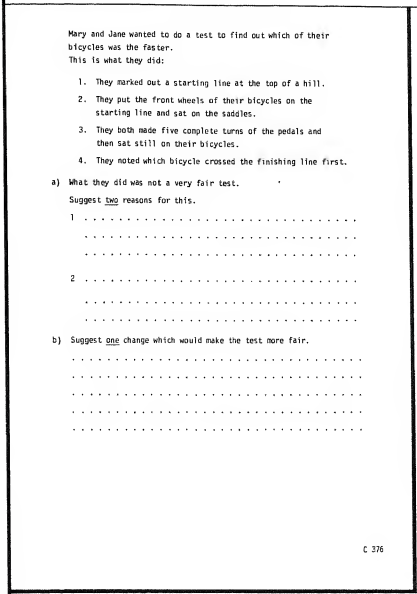
There were 6 questions of this type and mean scores ranged from 30% to 80%. Questions which required the use of science concepts had mean scores in the range 30% to 40%, questions that did not require taught science knowledge had mean scores 60% to 80%. The non-response rate was higher than for coded-answer questions, being 4% – 10% for concept-free questions and 12% – 19% for those requiring taught science concepts.

8.3.4 Criticising a procedure

These questions were similar to those described in 8.3.3 except that in this case the whole experimental procedure was given, and pupils had to suggest reasons for poor results, propose improvements, or say why the test had been "unfair". Example question 7 about bicycles is an example of this question type.

Criticising a procedure

Example question 7 (5β)



Mary and Jane wanted to do a test to find out which of their bicycles was the faster.
This is what they did:

1. They marked out a starting line at the top of a hill.
2. They put the front wheels of their bicycles on the starting line and sat on the saddles.
3. They both made five complete turns of the pedals and then sat still on their bicycles.
4. They noted which bicycle crossed the finishing line first.

a) What they did was not a very fair test.

Suggest two reasons for this.

1 . . . .

. . . .

. . . .

2 . . . .

. . . .

. . . .

b) Suggest one change which would make the test more fair.

. . . .

. . . .

. . . .

. . . .

. . . .

C 376

n= 765
Mean score 40%
Non-response 6%

Mark scheme

a. Any acceptable answer scores 1 mark, to a maximum of 2 marks.

Examples:	
The two girls may be different weights.	1
They may push off from the ground unequally	1
They push on pedals with different forces	1
One girl may be stronger than another when pedalling	1
Different shaped girls give different air resistance	1
Any two	2

b. Any acceptable answer scores 1 mark.

Examples:	
The same girl to test each bike and time the run	1
Each girl to test both bikes.	1
Use weights to get the same load	1
Not to pedal—freewheel downhill	1
Repeat the experiment.	1
Any one	1

Maximum score ③

Mapped score	Raw score	% pupils
3	3	13
2	2	28
1	1	26
0	0	33

In this question, many pupils wrote down things which could cause differences in performance, e.g. different sized wheels, different gears etc. instead of aspects of the test situation which should be controlled.

Questions of this type had lower mean scores than any other.

This question, based on an everyday situation, was one of the easiest, although the mean score is only 40%.

Boys scored 44% compared with girls 36%, a difference which was present in other questions of this type.

Performance on this group of questions was lower than on all the others with mean scores ranging from 20% – 60%. The effect of science concepts was again noticeable. For questions which depended on the use of science concepts the mean scores were 20% – 30% with non-response rates 13% – 25%, while for those that were concept free, the mean scores were 30% – 60% and non-response rates 5% – 12%.

In this group of nine questions the boys scored markedly better than the girls on five, with little difference in performance on the others.

8.3.5 Sequencing the steps in a procedure

In these questions pupils were given a procedure, with the steps listed in the wrong order, and they were asked to sequence them correctly.

This is illustrated in Example question 8, which was also used in the 1980 survey of 13 year olds (DES 1982)

Sequencing a procedure

Example question 8 (5β)

Bob has been asked to do a test to find out whether the pink colour in pink rose petals is a pure substance or a mixture.

He has been given this set of instructions for the test, but they are in the wrong order.

Put numbers 1 to 4 in the boxes to show the correct order.

(Put 1 in the box by the thing he should do first, 2 by the next thing, and so on.)

☐ A Crush sand, acetone and rose petals together.

☐ B Pour off the pink liquid into a beaker.

☐ C Add acetone, a drop at a time, to the middle of the filter paper.

☐ D Put a few drops of pink liquid in the middle of a filter paper.

C 7:3

This question shows the pattern characteristic of the type, that many pupils go wrong after the first step.

n =	765
Mean score	55%
Non-response	5%
Misinterpreted rubric	4%

Mark scheme

A – 1 $\boxed{1}$
B – 2 $\boxed{2}$
D – 3 $\boxed{4}$
C – 4 $\boxed{3}$

Start at A 1
Go to B 1
B to D 1
(D to C)

Maximum score ③

In all sequencing questions the final step is automatic, and so it gets no marks.

	% pupils age 15	% pupils age 13
1st stage correct	71	47
2nd stage correct	49	22
3rd and 4th stage	44	20

Raw score	% pupils age 15	% pupils age 13
3	44	20
2	5	2
1	22	25
0	29	52

Seven questions of this type were used, and this was the only group of questions in which the effect of science concepts did not seem to be important. Many of the questions, such as Example question 8, involved a standard laboratory technique or standard science procedure but pupils scored as well on those as on the more general ones. The mean scores ranged from 35% – 72%.

The hardest question of this type was a general one in which the pupils had to sequence the steps in the procedure to find the average length of a rice grain. This question included redundant steps, and despite the warnings in the rubric, 34% marked *every* box. The results on this question were consistent with those from Category 2 where most pupils measured only one bean when given 20 and asked to find the average length of one.

The other feature of these questions was the score distribution, in that very few pupils scored in the middle range: i.e. if they got past the first stage, they were then unlikely to go wrong, whether it was a 4 stage or 5 stage procedure. Those scoring zero did not start at the right step, and those scoring 1 started at the right step but then went wrong.

Between 1% – 4% of pupils misunderstood the question rubric, and put ticks in boxes or offered similar wrong responses. Few pupils failed to mark each box. Non-response rates ranged from 2% – 6%.

8.3.6 Selecting the statements which a given experimental procedure will test

Pupils were presented with an experimental procedure, and then a series of related statements. They had to relate the statements to the procedure, and mark boxes to show which statements could be tested by that experiment, and which could not. A few questions involved a multiple choice response but most had a series of statements and pupils had to assess each one, like true/false questions, as in Example question 9.

Selecting what an experiment will show

Example question 9 (5β)

A doctor decided to test two new kinds of sweets for treating sore throats.

He divided the patients with the same kind of throat infection into four groups.

The treatment for each group was:

Group 1 - took one sweet X every 4 hours
Group 2 - took one sweet X every 2 hours
Group 3 - took one sweet Y every 4 hours
Group 4 - took one ordinary boiled sweet every 4 hours

After two days the doctor examined the patients again.

Which of these questions do you think could be answered by this test?

Put a tick in the box by those that could be answered.
Put a cross in the box by those that could not be answered.
(Make sure you put a tick or a cross in every box.)

☐ A Does sweet X work better than a boiled sweet?

☐ B Does it matter how often sweet X is taken?

☐ C Does it matter how often sweet Y is taken?

☐ D Does sweet X have a different effect from sweet Y?

☐ E Does sweet Y work better than a boiled sweet?

C 23

n =	741
Mean score	56%
Non-response	4%
Partial response	5%

Mark scheme

Option		% pupils correct
A	✓	77
B	✓	64
C	×	63
D	✓	72
E	✓	74

Maximum score ⑤

Mapped score	Raw score	% pupils
3	5	31
2	4	30
1	3	15
0	2, 1, 0	24

Discussion

This question was about half-way along the facility range for this question type.

The mean score for girls was 59%, for boys 53%, a variation which was characteristic for the question type.

The distribution of the raw marks is common to this question type.

Girls scored marginally better than boys in six of the 11 questions of this type. As with most other question types, those questions involving a science concept had lower mean scores. The non-response rates ranged from 0% – 6% and between 4% – 6% of pupils did not follow the rubric correctly, failing to mark all the boxes.

8.4 Summary of pupil performance on sub-categories 5α and 5β

24 questions were used to represent sub-category 5α, *assessing testable statements,* with eight in each of three test packages. About 4% of pupils scored no marks, and around 10% scored full marks. The distribution of marks was skewed slightly towards the higher scores, giving a mean score of about 60% for this sub-category.

In sub-category 5β, *experimental procedures,* there was a greater variety of question types as already discussed, and fifty-four questions were used with eighteen in each of the three test packages. About 5% of pupils scored 10% or less, and about 10% of pupils scored over 70%. Of the 2,200 pupils tested, only 19 (less than 1%) scored over 90%, and only one pupil scored full marks on one sub-test package. The distribution of marks was fairly symmetrical, giving a mean score of around 50% for this sub-category.

There was no difference between the overall performance of boys and girls on either sub-category. However there was some variation on individual questions and question types. Boys were slightly better on questions requiring criticism of experimental procedures, girls were slightly better on those requiring the selection of the statements an experiment would test. These findings agree in principle with those from Category 4 that although boys overall have a better knowledge of physical phenomena, girls appear to be better at questions needing analysis.

Non-response rates varied from 0–25%, being lowest on multiple-choice questions, slightly higher on coded-answer questions when two or more boxes needed marking or when questions required the use of science concepts, and highest where an extended response was required to a question involving science concepts.

Various misinterpretations or misreadings of the rubric occurred; the most noticeable being partial responses when more than one box needed to be marked (up to 7%); and up to 4% of the pupils did not appear to understand what to do on the sequencing questions.

Many questions in the 5β sub-category, *assessing experimental procedures*, did not require the pupils to have any detailed knowledge of science or science concepts. All necessary information was given in the questions, or could be expected from day to day experience, and all that was required was the application of some general principles, such as control of variables. Variations in the context in which questions were set seemed to have an effect on the scores. Those questions that appeared "scientific", being set in a laboratory or using laboratory apparatus, in general were not answered as well as those requiring the application of similar principles in an everyday setting.

Certain of the category 5 questions were used at age 13 and at age 15, and some of these have been discussed. While comparisons of performance can be made at this individual question level, it is not possible to compare performance on question types or overall sub-category scores, since the contexts and amounts of taught science knowledge needed was different, and different numbers of questions were used.

8.5 Devising and describing investigations: sub-category 5γ

8.5.1 Introduction

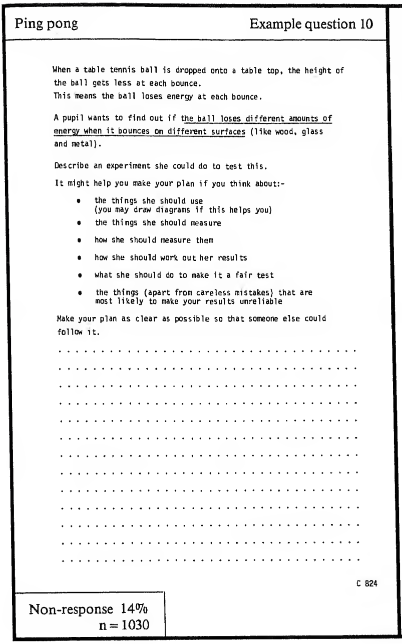
In these questions pupils were asked to give written accounts of the experiments they would do to investigate a given problem or to test a given hypothesis. The answers entailed more writing than most others asked for in the survey, each occupying pupils for up to 15 minutes.

5 questions of this type were used in the 1980 survey, each being given to about 1,000 pupils. Two of these questions are reported in detail here; results from these and the other γ questions will be incorporated with results from the individual practical investigations from the 1981 survey, to report in more detail on pupils' performance on investigations.

All the questions had a complex grid for the marker together with coding instructions for filling in the grid. The completed grid gave a description of pupils' performance in relation to different aspects of experimental design. Features of pupils' responses which were coded for each question included the general strategy adopted, the control of variables, and how they could be controlled, a description of the measurements to be made and how they could be made, the way the results are analysed and an account of possible errors or an indication of the need for repetition of measurements.

From the features coded for each question, four have been isolated which are essential for a successful design. These are identifying the independent variable, measurement of the dependent variable, the control of other variables, and the use of results.

Two questions, Example questions 10 and 11, are described in detail here; for each one the proportion of pupils incorporating each of the features in their design was investigated, followed by an analysis of the proportion of pupils using combinations of the features.



Ping pong **Example question 10**

When a table tennis ball is dropped onto a table top, the height of the ball gets less at each bounce.
This means the ball loses energy at each bounce.

A pupil wants to find out if the ball loses different amounts of energy when it bounces on different surfaces (like wood, glass and metal).

Describe an experiment she could do to test this.

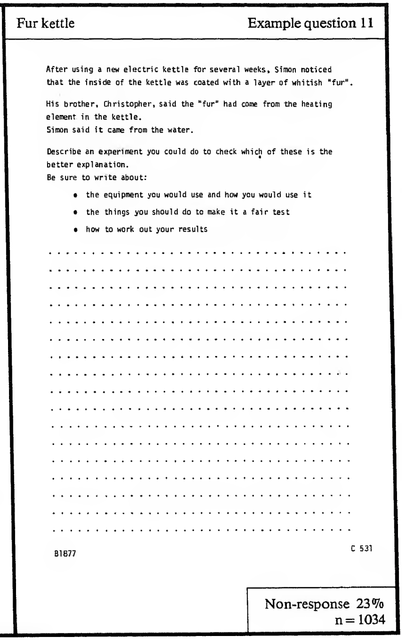
It might help you make your plan if you think about:-

- the things she should use (you may draw diagrams if this helps you)
- the things she should measure
- how she should measure them
- how she should work out her results
- what she should do to make it a fair test
- the things (apart from careless mistakes) that are most likely to make your results unreliable

Make your plan as clear as possible so that someone else could follow it.

C 824

Non-response 14%
n = 1030



Fur kettle **Example question 11**

After using a new electric kettle for several weeks, Simon noticed that the inside of the kettle was coated with a layer of whitish "fur".

His brother, Christopher, said the "fur" had come from the heating element in the kettle.
Simon said it came from the water.

Describe an experiment you could do to check which of these is the better explanation.
Be sure to write about:

- the equipment you would use and how you would use it
- the things you should do to make it a fair test
- how to work out your results

B1877 C 531

Non-response 23%
n = 1034

8.5.2 Ping pong

In this question, pupils were asked to devise an experiment to find out on which surface a bouncing ball loses most energy. The vast majority of pupils suggested one of two methods, although 14 pupils suggested totally different methods.

Table 8.1 *Methods suggested by pupils*

	% pupils
Rebound height only	50
Number of bounces only	11
Both methods above	7
Failed to start adequately	32

As Table 8.1 shows the most popular experimental method chosen to investigate the problem was to measure the rebound height of the ball from the different surfaces; half of the pupils offered this approach. A more infrequent suggestion was to count the number of bounces of the ball on the different surfaces. About one pupil in three failed to make a satisfactory start to the question. These unsuccessful pupils were about evenly divided into those who made no attempt at all, and those who described totally inadequate methods or appeared not to understand the task.

Some features of the experimental design are common to both main methods—the different surfaces must be used, and the initial height of the ball must be controlled. 40% of the pupils stated the necessity to control the initial height, and a further 15% implied that this needed doing. Over half of the former group gave details of how they would attempt to exert such control. 46% of the pupils specified the use of different surfaces to bounce the ball on and a further 13% implied this use. Those who failed with these conditions wrote very general statements like "Bounce the ball and watch what happens".

Diagrams were drawn by 17% of the pupils, but 6% of these were at best unhelpful and at worst confusing.

About one in ten pupils said that they would use surfaces of the same size—a variable admittedly, but in this case one which is irrelevant to the outcome of the experiment. Other controls, such as using the same ball, were mentioned by 30% of pupils.

15% of the pupils said that the results should be checked by repetition. This is higher than found on other questions; maybe because the pupils can envisage problems such as the ball bouncing off surfaces. The repetition would then be necessary to achieve a result at all, rather than to improve it.

Rebound height method

This was the method chosen by 57% of pupils.

The main features of the method are summarised below.

Table 8.2 *Features of "rebound height" method*

	% pupils	% of those using method
Rebound height method	57	100
1. using different surfaces	44	77
2. noting the rebound height	38	67
3. controlling the initial height	41	72
4. use of results	19	33

Of those who said they would measure the initial rebound height, just less than half also said they would note the subsequent rebound heights, and just over half explained how they would try to measure these heights. Some of the ideas were very ingenious, like having coloured strings at fixed height intervals.

Of those who attempted to indicate how they would use their results, about a third made a correct deduction, for example "The higher the bounce or greater number of bounces the less the loss of energy", and another third gave a very vague statement like "Look at the results and see".

Possible errors were mentioned by 16% of the pupils, and although only 1% specifically mentioned the word "parallax", 9% mentioned the difficulty of judging the height of the ball.

Combinations of features

While it is interesting to note how many pupils mention the individual features, from the point of view of a successful design it is important to know how many combine the features to make up a sensible protocol. For each experiment there were almost as many different combinations as pupils. These have been rationalised into a logical sequence, starting with the general method adopted, and following on with identifying the independent variable, measuring the dependent variable, controlling the initial conditions, and use of the results.

When looking at the tables on combinations of features, it is important to realise that the percentages are inclusive: the 43% who say they would use the different surfaces *includes* the 30% who go on to so they would measure the rebound height. By difference, 13% who say they would use the different surfaces do *not* mention that they would measure the rebound height.

Comparing the figures on the combination (eg. from Table 8.3 with the figures for those who included the isolated feature (eg. from Table 8.2) shows the number who failed to include previous features in their design.

Table 8.3 *Combinations on "rebound height" method*

Rebound method	✓	✓	✓	✓	✓
1. different surfaces		✓	✓	✓	✓
2. rebound height			✓	✓	✓
3. control initial height				✓	✓
4. use of results					✓
% Total	57	43	30	22	9
% on this method	100	75	53	40	16

What the combinations show clearly is that while 30% – 40% of the pupils mentioned the main design features, (ie. half of those who choose this method do something sensible), the proportion who can follow through to what they would do with the results falls considerably.

Number of bounces method

This method was chosen by 18% of the sample. There were several variations on this general method, and these included suggestions to count the total number of bounces, to note the time to stop bouncing, to measure the time for a fixed number of bounces or the number of bounces in a fixed time. The great majority of the pupils opting for this general method offered one or other of the first two possibilities.

The basic features of the method are shown below.

Table 8.4 *Features of "number of bounces" method*

	% of pupils	% on method
Number of bounces method	18	100
1. using the different surfaces	14	78
2. counting the number of bounces	14	78
3. controlling the initial height	13	72
4. use of results	5	28

Of the 5% attempting to describe how they would use their results, half gave a detailed and correct account, the rest were muddled or unspecific.

Possible errors were mentioned by 10 of the 174 pupils who chose this method, only two pupils mentioning the problem of distinguishing and counting the bounces towards the end.

As with the first method omitting to indicate how an answer could be worked out from the results was the major failure point.

Combination of features

When looking at the combinations on this method, it is again apparent that while most pupils manage to get some of the points, few manage to put a whole procedure together.

Table 8.5 *Combinations on "number of bounces" method*

Number of bounces method	✓	✓	✓	✓	✓
different surfaces used		✓	✓	✓	✓
no bounces counted			✓	✓	✓
control initial height				✓	✓
use of results					✓
% Total using this combination	18	13	11	8	3
% those on this method using this combination	100	72	61	44	17

8.5.3 Fur kettle

This question is slightly different in style from "Ping pong"; in it pupils are given a situation and two hypotheses that have been put forward to explain it. They are asked to design an experiment to find out which hypothesis is correct.

As with "ping pong", two general methods were adopted: *either* keeping the element the same and changing the water, *or* keeping the water the same and removing the element. Of the 12% of pupils who suggested a completely different method, all but a few were judged by markers to be inappropriate, with the pupils concerned being unable to come to a conclusion.

Table 8.6 *Pupils' suggested methods*

	% pupils
Removing element only (water same)	17
Changing water only (element same)	22
Both methods above	6
Never started adequately	41
Other methods	12

Of the 41% of the pupils who failed to start satisfactorily, over half made no attempt to answer the question, and most of the remainder offered totally inadequate methods.

In their answers to this question, many pupils showed that they knew the fur came from the water, and 4% gave this as their answer without attempting to describe an experiment. However, their knowledge of what the fur was and how it came about was very confused. In the event such knowledge did not help the pupils with this question and 9% of pupils confused their answers by trying to use their "knowledge" of hard water.

In each method there are controls that are necessary to get a valid result. However, 7% of pupils mentioned other non-relevant variables, such as the volume of water used or the number of times the water was boiled. In general, these pupils did *not* mention controlling the variable fundamental to their experiment—ie. the element or the water.

Changing the element method (water the same)

23% of the pupils chose this method, which can be broken down into the following component parts:

Table 8.7 *Features of "changing the element" method*

		% of pupils	% on method
	Change element method	23	100
1.	remove the element	22	96
2.	look for fur	17	74
3.	use water same source	10	43
4.	use of results	12	52

The majority of those who mentioned the necessity of removing the element also explained how they would do it. Of those attempting an analysis, the majority concluded correctly, very few (less than 1%) using the vague "see who is right" form.

Combinations of features

Again the results show that few pupils were able to put the whole procedure together.

Table 8.8 *Combinations on "changing the element" method*

Change element method	✓	✓	✓	✓	✓
1. remove element		✓	✓	✓	✓
2. look for fur			✓	✓	✓
3. use water same source				✓	✓
4. use of results					✓
% total	23	22	16	8	4
% on this method	100	96	70	35	17

Although using water from the same source is an *essential* control to getting a result in this experiment, many pupils omitted to mention this and instead gave vague comments like "take the element out and look if the fur is still there".

Changing the water method (element same)

This method proved to be the more popular of the two, 28% of pupils choosing to use. It can be broken down into the features shown below.

Table 8.9 *Features of "changing the water" method*

	% of pupils	% on method
Changing water method	28	100
1. use different water/liquid/none	27	96
2. look for fur	17	61
3. keep element the same	10	36
4. use of results	16	57

All the pupils using this method mentioned using different water in some way or another, about a third of them suggesting that no liquid be used at all! More than half of those who said they would use water from a different source, or a different liquid, said what they would use; choices ranged from distilled water to milk, oil, ink . . . !

The vast majority of those who analysed their results did so correctly.

Table 8.10 *Combinations on "changing the water" method*

Change water method	✓	✓	✓	✓	✓
1. different liquid		✓	✓	✓	✓
2. look for fur			✓	✓	✓
3. element same				✓	✓
4. use of results					✓
% total	28	27	18	6	4
% on this method	100	96	64	21	14

Again there were two features commonly omitted in pupils' descriptions: the need to control the variable and specifying how they could work out an answer from their results.

8.6 Discussion of performance on "Design of Investigations"

In view of the finding from the HMI secondary survey (DES 1979) that in only about half of schools did pupils have any opportunity to design experiments themselves, it might be thought that the performance levels on the first two sub-categories are higher than might have been expected. However, in these two sub-categories it is only the isolated, component, parts of the overall skill of experimental design that are tested; these have to be brought together in the γ group of questions when a complete experimental description is asked for. In these questions about a third of the pupils did not manage to suggest a sensible method; of those who did start, about half gave very short, general responses without any detail. The rest managed to cover the main points of the experiment, but in a random manner, so that very few put together a coherent sequence. Of the pupils who did describe an appropriate procedure many omitted to mention the need to control particular variables and also failed to indicate how the results could be used to work out an answer to the problem posed. Pupils can be descriptive in the method, but then fail to be analytical. Although not marked on presentation, markers commented on the general lack of structure of the writing. The haphazard way the points were presented indicated that often little planning or thought had taken place before writing had begun.

In general, about two-thirds of the pupils can pick out an untestable statement and can rephrase, about half can select the appropriate procedure, sequence a procedure, and know what an experiment will test. When asked to design a whole investigation about one in ten pupils manage to include all the basic features necessary for a successful experiment.

9 Pupils' science performances

9.1 Introduction

Throughout Chapters 3 to 8 of this report individual questions have been presented and pupils' responses to these questions have been discussed. Occasionally, generalisations have been made about pupils' performances on groups of questions of similar type and comments offered about overall performance levels on the various sub-categories which form the assessment framework.

This chapter considers overall sub-category performances in greater detail, and in particular compares the performances across the complete profile of sub-categories for various pupil sub-groups (for example, boys and girls, pupils in metropolitan and non-metropolitan locations, and so on). Readers are reminded that "test scores" were not produced for every sub-category. This has been for educational reasons: often because the questions concerned were very specific and directly interpretable (for example, the questions of sub-category 3α, *using an identification key,* and those of sub-category 1γ, *using scientific symbols and conventions*); sometimes because the skill concerned was not considered of such major importance as to justify a large amount of survey testing time.

The sub-categories to be discussed in this chapter are:

1α	Reading information from graphs, tables and charts
1β	Expressing information as graphs, tables and charts
2α	Using measuring instruments
3β	Observing similarities and differences
4α	Describing and using patterns in information
$4\delta_{AB}$	Applying biology concepts
$4\delta_{CD}$	Applying physics concepts
$4\delta_{EF}$	Applying chemistry concepts
5α	Identifying or proposing testable statements
5β	Assessing experimental procedures

Sub-categories 1α and 1β are complementary to each other—both concern data presentation skills necessary for work in science. Sub-category 4α depends in some part on similar skills, because pupils here have to *interpret* information gathered from some form of data presentation. Practical science skills are involved in sub-categories 2α and 3β: both were assessed by tests incorporating practical work in a laboratory setting. 5α and 5β concern basic design skills.

Sub-category 4δ differs considerably from the other sub-categories because performance here requires pupils to apply their knowledge of facts and concepts of science commonly taught in schools. As outlined in Chapter 7, three different 4δ sub-categories have been identified: $4\delta_{AB}$ which requires the use of concepts of biology; $4\delta_{CD}$ which requires the use of concepts of physics; and $4\delta_{EF}$ which requires the use of concepts of chemistry.

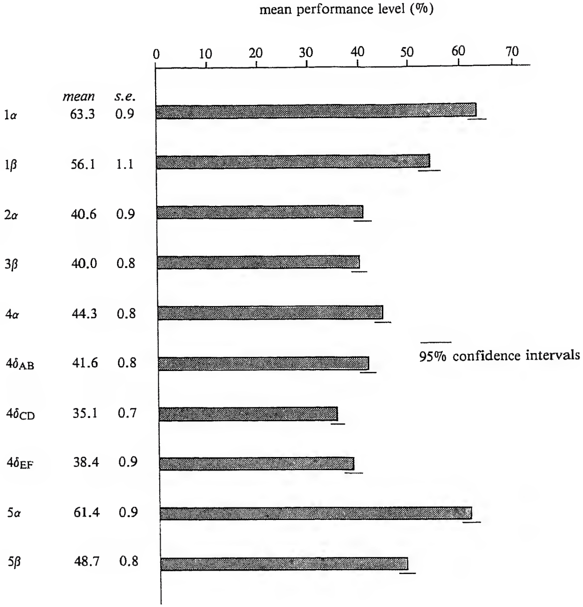
Performance on all of these sub-categories will be considered in relation to a number of general background variables (including pupil sex, region, location, school type and size, pupil-teacher ratio and expenditure on teaching resources), and also in relation to a number of science-specific pupil and school characteristics (including curriculum background, science expenditure, laboratory accommodation and technician help). A summary table of results is presented in Appendix 7.

The performance figures presented in this chapter are all *estimates* of the levels of performance of the respective pupil sub-groups. These estimates are based on the performances of pupils contained in the survey sample. The accuracy of a performance level estimate depends essentially on the size of the pupil sample on which the estimate is based, on the size of the pupil group taking the same test in each school, and on the variability of sub-category performance among these pupils and schools. The standard errors associated with the performance estimates are given in each graphical presentation, as are the 95% confidence intervals. Comments are also occasionally made about the statistical significance (at the 5% level) of particular sub-group performance differences. A visual rule of thumb indication of statistical significance is given by the confidence intervals: where these do not overlap, or overlap only slightly, then the performance levels concerned are likely to differ significantly from one another in the statistical sense for that particular sub-category. (Readers unfamiliar with the terms standard error, confidence interval and statistical significance are advised to refer to Appendix 8).

Figure 9.1 shows estimates of 15 year old pupils' overall performance levels on the ten sub-categories.

While it is tempting to make direct comparisons of these levels from one sub-category to another, it is very difficult to assess the educational significance of these performance differences; this issue is discussed in the next chapter. The vertical lines on the figure (the 95% confidence intervals) indicate the extent to which the estimated levels are in doubt. There is a 95% probability that the true level lies within the range indicated.

Figure 9.1 *Overall sub-category performance levels (all 15 year old pupils in England, Wales and Northern Ireland)*



9.2 Pupil performance in relation to general school characteristics

Pupils can be classified in a number of ways by virtue of the fact that they attend particular kinds of schools. For example, the pupil sample can be divided into those pupils attending schools in the North of England, those attending schools in Wales, and so on. Alternatively, those pupils attending 11 – 18 comprehensive schools can be distinguished from those pupils attending 11 – 16 comprehensives. Some of these school variables were considered in Chapter 2 of this report when the survey sample was described.

Although a survey such as this, with its many sub-categories and numerous different test packages, has a more limited scope for the investigation of relationships between performance and background variables than would a survey of this size involving a single kind of test, some investigation of background variables was nevertheless possible.

This first section considers those descriptive variables which are not specific to science education in the schools. These include type of school, region, location in England, and pupil-teacher ratio. More science-specific characteristics such as laboratory provision, science expenditure and curriculum background are considered later.

The search for links between some of these variables and pupils' science performances was often restricted to pupils in comprehensive schools. This was to avoid masking (or falsely suggesting) a relationship between performance and a particular school variable by, for instance, comparing two or more groups one of which contained more pupils from, say, modern schools (and hence contained a different ability range) than the other.

9.2.1 Performance by region and school location

Figure 9.2 shows that pupils in comprehensive schools in non-metropolitan locations consistently achieve higher estimated performance levels than pupils in metropolitan comprehensive schools (only the 3β sub-category performance difference reaches statistical significance, however, and this is indicated in the figure by the fact that the 95% confidence intervals for the metropolitan and non-metropolitan estimates for 3β do not overlap).

Figure 9.3 illustrates the regional patterns of estimated sub-category performance. A high degree of consistency is apparent in this figure: performance estimates for pupils in the South and Midlands being at a generally higher level than those for pupils in the North, and in Wales and Northern Ireland.

These regional and location findings in England have already been established for 11 and 13 year olds in the APU science surveys at these ages (DES 1981a, 1982a). They overlap to some extent because, as noted in Chapter 2, the North region has proportionally more schools in metropolitan counties than have the Midlands and South.

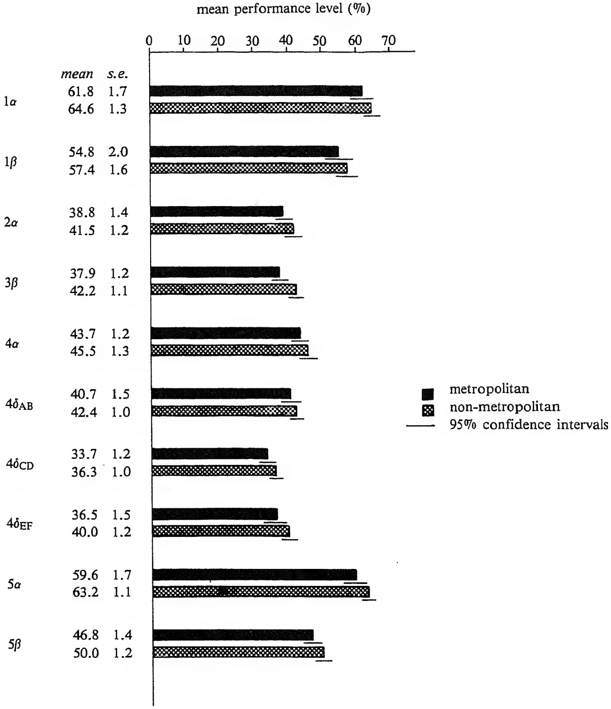
Within England it has been suggested that regional differences in pupils' attainments are linked to differences in the social class composition of these regions, given that proportionally more middle class people live in the South than in the Midlands and more in the Midlands than in the North (Davie et al. 1972). This possibility is supported by the large regional differences discussed later in this chapter with regard to levels of uptake of free school meals (an often-used surrogate socioeconomic index). The relative affluence of the South in particular was also reflected in the higher levels of resource provision in the survey schools in this region.

9.2.2 Performance by type and size of school

There were no statistically significant differences in the estimated sub-category performances of pupils in the three kinds of comprehensive school distinguished by age-range (11 – 16, 11 – 18 and 13 – 18). Neither was there any evidence of a relationship between *size* of school and pupils' science performances.

Survey pupils attending comprehensive schools were classified by the size of their schools, and estimated performance levels were calculated for three size ranges: *1 – 600, 601 – 1,000*, and *1001 or more pupils* on roll. There were no

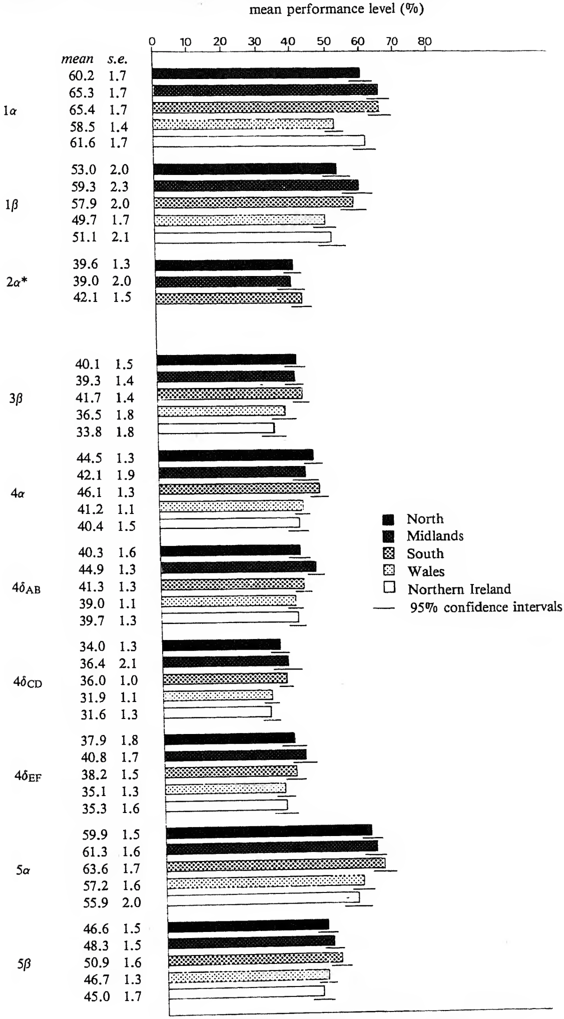
Figure 9.2 *Sub-category performance by school location (15 year old pupils in England)*



statistically significant performance differences for any one sub-category, nor any indication of a consistent pattern.

These results are in line with those reported for the previous age 13 science survey (DES 1982a). No relationship could be established between 13 year old pupils' science performances and the size of their comprehensive schools, nor were there any differences between the performance levels of pupils in comprehensive schools with a leaving age of 12/13 years compared with those of pupils in comprehensive schools with higher leaving ages (ie. 16 or 18).

Figure 9.3 *Regional patterns of sub-category performance*



* 2α was not assessed in Wales or Northern Ireland in this first survey.

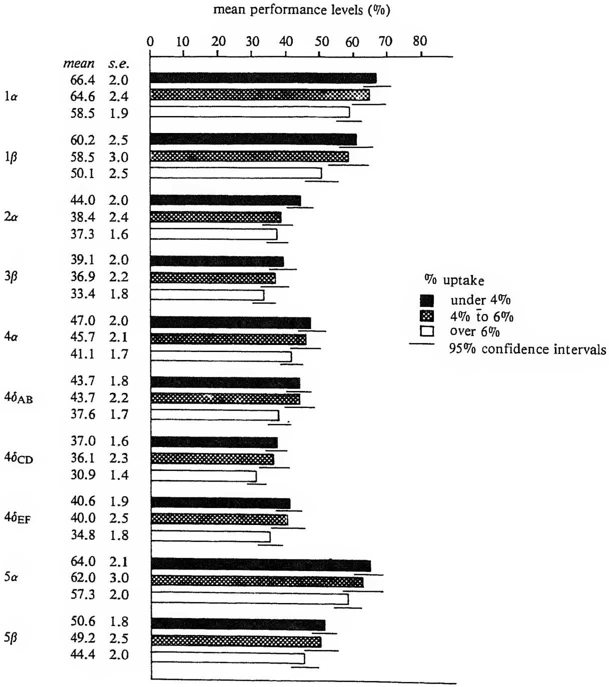
9.2.3 Percentage uptake of free school meals

Percentage uptake of free school meals has often been found to be a useful socioeconomic indicator, strongly associated with pupils' academic attainment (see, for example, Gray 1981).

In this survey, every maintained sample school was asked to record the number of pupils taking a free school meal on a specified day near the testing period, so that the percentage of pupils on roll taking a free school meal could be computed to serve as an index of the relative affluence of that school's catchment area.

In order to maintain consistency with the previous report for age 13 pupils (DES 1982a), performance estimates were produced for schools classified into three groups based on the following ranges of percentage uptake of free school meals: under 4%, 4% to 6%, over 6%. Figure 9.4 presents the results of this

Figure 9.4 *Percentage free school meals uptake and sub-category performance (pupils in comprehensive schools in England)*



exercise, a clearly consistent pattern of higher performance being associated with lower percentage uptake.

As intimated previously, there is a high degree of overlap between this finding and the regional and location patterns of performance described earlier. About 60% of the survey schools in the North, Wales and Northern Ireland had free school meal uptakes higher than 6%, compared with 25% of the survey schools in the Midlands and South. Also a higher proportion of the metropolitan schools fell into the high uptake group than did those English survey schools in the non-metropolitan locations (43% and 29% respectively).

9.2.4 Pupil/teacher ratio

There was no evidence in the survey results of any relationship between pupils' science performances and overall pupil/teacher ratio.

Chapter 2 described how the survey schools fell into three major groups distinguished by pupil/teacher ratio: *1 – 14.9*, *15 – 17.4* and *17.5+*. Performance estimates were computed for these three groups for comprehensive school pupils only. There were no statistically significant differences between the estimated performance levels of the groups for any sub-category.

There are, of course, variations in the way in which different LEAs compute pupil/teacher ratio statistics. Headteachers and their deputies are usually included in the computation. Schools with sixth forms will also tend to have lower overall ratios. Pupil/teacher ratios cannot, then, be assumed to relate in any clear way to the sizes of teaching groups in the schools concerned (especially for science), so that this particular result should not be assumed to convey any information about the way in which pupil performance and size of teaching group might be associated.

9.3

Pupil performance in relation to science-specific school characteristics

A number of science-specific school characteristics were considered in Chapter 2 when the survey school sample was described. These include the proportion of science-qualified teaching staff in the schools, laboratory provision, and school expenditure on science.

In that chapter it was noted that, overall, 95% of the staff who were teaching science in the survey schools held a science qualification (science or applied science degree, or certificate course with science a main subject). There was insufficient variation from school to school in terms of this variable for investigation of association with science performance to be possible.

In this section, then, we consider the three variables of adequacy of laboratory accommodation, adequacy of technician support and level of annual expenditure on science education.

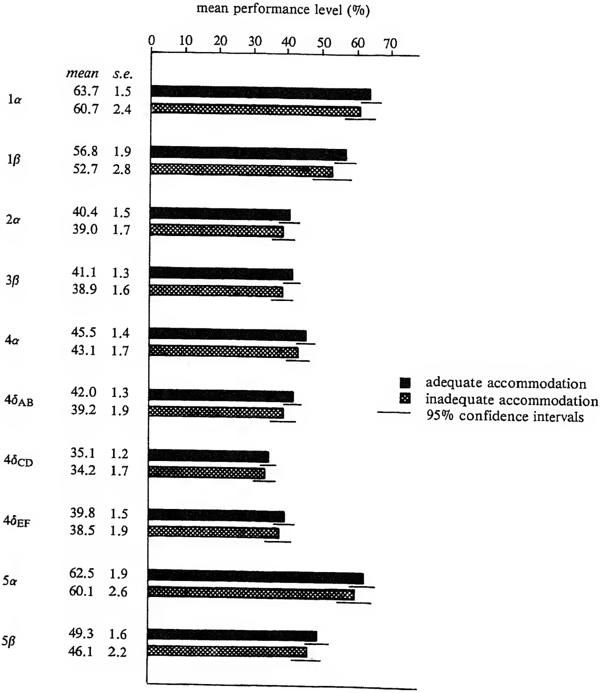
For the reason given previously, these investigations were confined to pupils in the comprehensive sector in England.

9.3.1 Laboratory accommodation

As described in Chapter 2, HMI's criterion for judging the adequacy of a school's laboratory accommodation was applied to each of the survey schools (DES 1979). HMI based their criterion on the assumption that the science courses of pupils over 11 years of age would be practically-based, and that such pupils would study science for about one-sixth of their total curriculum time.

Performance estimates were produced for each sub-category for pupils in each group of schools, ie. for pupils in schools *with* adequate accommodation, and for pupils in schools *without* adequate laboratory accommodation. The resulting performance profiles are shown in Figure 9.5.

Figure 9.5 *Laboratory accommodation and sub-category performances (15 year old pupils in comprehensive schools in England)*

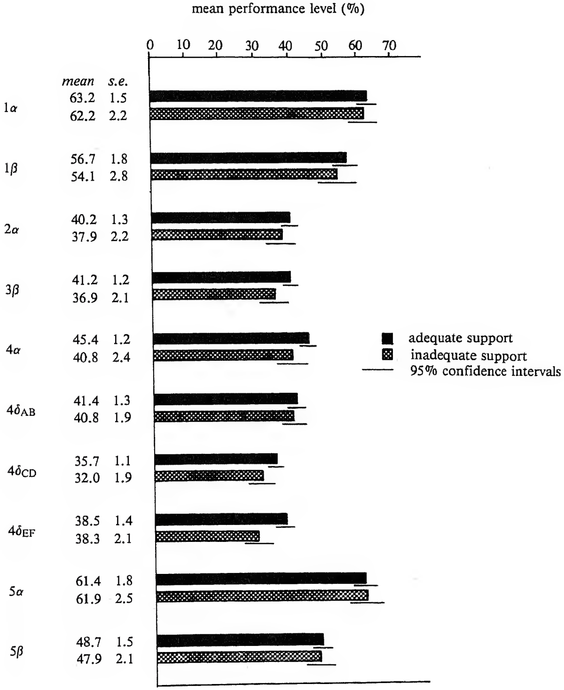


Although none of the sub-category performance differences reaches statistical significance, it is nevertheless clear from Figure 9.5 that pupils in schools with adequate laboratory accommodation tend to achieve consistently higher performance levels than those pupils in schools with insufficient numbers of laboratories according to the HMI criterion.

9.3.2 Technician support

HMI judged a school's technician support to be adequate if each full-time equivalent technician was required to service three or fewer laboratories (DES 1979). On the basis of this criterion, Figure 9.6 shows the sub-category performance estimates for pupils in English comprehensive schools with adequate technician support for their laboratories and for pupils in comprehensive schools with insufficient numbers of technicians for their needs.

Figure 9.6 *Technician support and sub-category performances (15 year old pupils in comprehensive schools in England)*



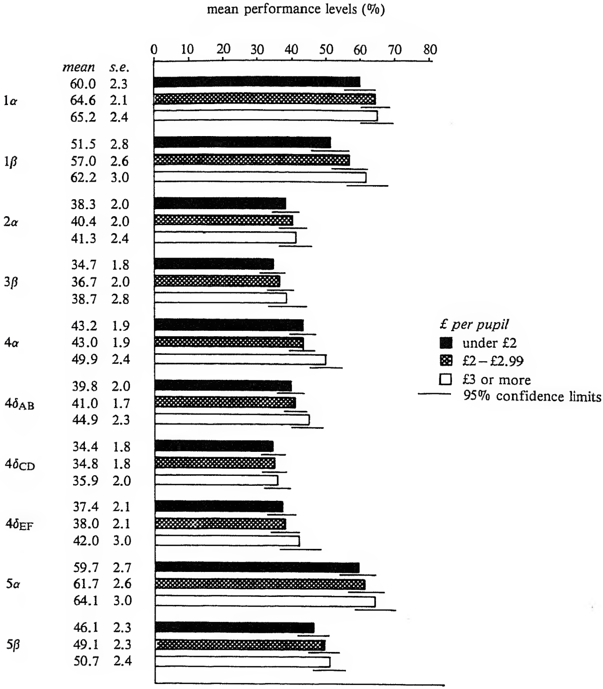
As was the case when laboratory accommodation was considered, none of the sub-category performance differences shown in Figure 9.6 reaches statistical significance, although there is again evidence of a consistency in the way in which higher pupil performance is associated with adequate technician support.

9.3.3 Annual science expenditure

Sub-category performance estimates were calculated for pupils in comprehensive schools spending *under £2, £2 – 2.99* and *£3 or more* on science equipment and other science teaching materials.

From the consistency in the directions of performance differences evident in Figure 9.7, it would appear that pupils' science performances are indeed

Figure 9.7 *Schools' average annual science expenditure and sub-category performances (15 year old pupils in comprehensive schools in England)*



	mean	s.e.
1α	60.0	2.3
	64.6	2.1
	65.2	2.4
1β	51.5	2.8
	57.0	2.6
	62.2	3.0
2α	38.3	2.0
	40.4	2.0
	41.3	2.4
3β	34.7	1.8
	36.7	2.0
	38.7	2.8
4α	43.2	1.9
	43.0	1.9
	49.9	2.4
4δ$_{AB}$	39.8	2.0
	41.0	1.7
	44.9	2.3
4δ$_{CD}$	34.4	1.8
	34.8	1.8
	35.9	2.0
4δ$_{EF}$	37.4	2.1
	38.0	2.1
	42.0	3.0
5α	59.7	2.7
	61.7	2.6
	64.1	3.0
5β	46.1	2.3
	49.1	2.3
	50.7	2.4

associated with school levels of expenditure on science education. As might be expected, the higher the level of science expenditure the higher the levels of performance of the pupils.

It is pertinent to note at this point that no association was found between pupils' science performance levels and schools' general level of expenditure on teaching resources. A similar lack of association was found for the science performances of age 13 pupils, and has previously been noted by other researchers for performance in other curriculum areas (see, for example, Rutter *et al.* 1979).

9.4 Sex and curriculum background

The estimated performance profiles for boys and girls are illustrated in Figure 9.8.

The performance differences for 1α, 2α, 4α and $4\delta_{CD}$ are all statistically significant and are all in favour of boys. While not reaching statistical significance, the large performance difference for 1β is also in favour of boys.

The sex difference in applying physics concepts was found also for 11 year old and 13 year old pupils in the 1980 APU surveys (DES 1981a, DES 1982a), and has previously been found in science surveys in other countries. For example, the IEA survey (Comber and Keeves 1973), the United States NAEP surveys (cf. NAEP 1978a, 1978b), and the British Columbia Science Assessment (Hobbs *et al* 1978) all found significant differences in favour of boys on their content-based tests involving physics. The evidence from these surveys indicates that this difference has already appeared by 11 years of age, and suggests that it increases with age.

The NAEP surveys also found boys to be significantly more competent than girls at all ages in the use of apparatus and measuring instruments (our sub-category 2α), and there was some evidence to this effect in the APU science surveys of 11 and of 13 year olds previously referred to. In fact, a special investigation of practical skills conducted in the United States (NAEP 1975) found that the difference between boys and girls was already apparent by the age of 9.

It must be of educational significance that boys and girls show no overall performance differences on the sub-categories concerned with observation skills (3β) and basic design skills (5α and 5β). The British Columbia Science Assessment noted this same feature of pupils' science performances. It may be, as discussed in the next chapter, that the performance differences that *have* been found can to some extent be accounted for by the different curriculum experiences of boys and girls.

It is very difficult to assess formally the influence of different curriculum experiences on pupils' levels of performance on these sub-categories. This is because there are a number of interacting variables at play—the three most influential ones being sex, ability and science subjects studied—and it is not possible to isolate the effect of any one of these without a specially designed investigation.

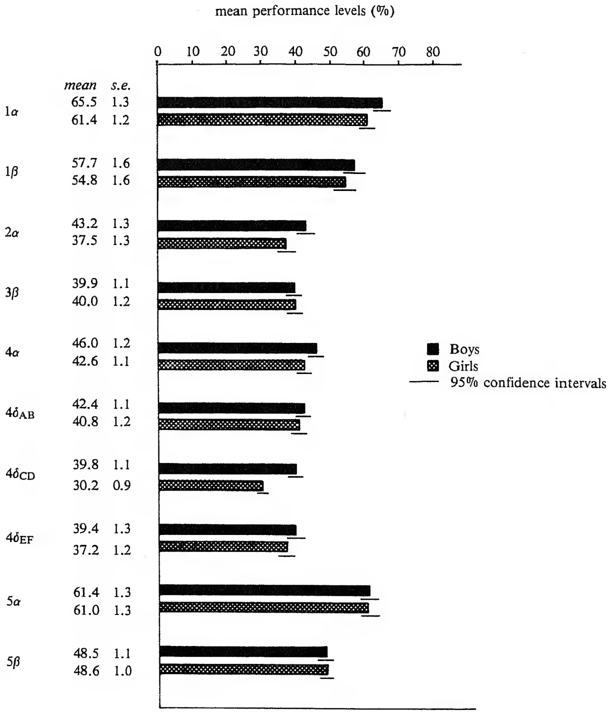
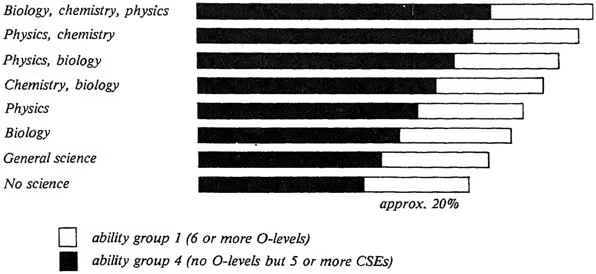


Figure 9.8 *Sex differences in sub-category performance (15 year old pupils in England, Wales and Northern Ireland)*

In addition to the sex difference in curriculum experience already referred to, it was noted in Chapter 2 that there are also differences in the kinds of science subjects being studied by pupils of different academic abilities (as indicated by the number and nature of external examinations for which the pupils were to be entered the following spring). In particular, pupils taking combinations of biology, chemistry and physics tended to be potential O-level candidates (most of those taking all three of these subjects being classified into the highest "ability band" by their teachers), while pupils taking just one of these subjects, human biology or general science were probable CSE only candidates.

Figure 9.9 *Curriculum group differences in sub-category performance: generalised pattern*



In order to make sensible comparisons of the performances of pupils taking these various subjects it would be necessary to take account of their different abilities. As the majority of sample pupils had been classified by their schools into group 1 (6 or more O-levels with or without CSEs) or group 4 (no O-levels but 5 or more CSEs) curriculum background comparisons were made within each of these pupil groups. However, sample sizes were still often too low to justify recording actual performance estimate figures (for example, the number of pupils in group 1 who were not taking any science was very low for every sub-category, as was the number of group 4 pupils taking biology, chemistry and physics as separate subject courses).

Similar patterns of performance did emerge, however, for the pupils in these two groups, and Figure 9.9 presents a rough indication of what this fairly consistent pattern was. It can be said that performance levels tended to be higher the greater the number of science subjects studied. This result cannot be a surprising one since the amount of curriculum time spent by pupils on science will be directly proportional to the number of science subjects studied, and, further, the greater the number of subjects taken the broader a pupil's science knowledge base will be.

There was clear evidence that the content of particular science subject courses influenced performance on related sub-categories. For example, pupils taking physics, particularly in combination with chemistry, seemed to have an advantage over others on 1α and 1β (graphical skills), and most clearly on $4\delta_{CD}$ (applying physics concepts). Pupils taking chemistry in any combination achieved higher levels than all others on $4\delta_{EF}$ (applying chemical concepts), while biologists achieved relatively higher performances on 3β (observation skills) and $4\delta_{AB}$ (applying biology concepts) than they did on other sub-categories.

Pupils taking a single physics or biology course, consistently performed at higher levels than pupils taking no science at all, the physics only pupils achieving performance levels above the biology only pupils on all sub-categories. General science pupils performed at only marginally higher levels than pupils taking no science.

9.5 Summary

There were consistent regional and school location performance differences across all sub-categories. Pupils in the Midlands and, particularly, the South achieved higher performance levels than those in the North, Wales and Northern Ireland, and pupils in schools in non-metropolitan areas in England achieved consistently higher performance levels than those pupils in schools in metropolitan locations. Such regional differences within England have been linked to the different social class distributions of the regions, and indeed this possibility is supported by the differential regional and location patterns of percentage uptake of free school meals in the survey schools, given that a strong association was found between pupils' levels of science performance and this variable.

No evidence was found of any association between the levels of performance of pupils in Comprehensive schools in England and *either* the age-range *or* the size of their schools.

Neither was there any suggestion in the data of any link between pupils' performance levels and their schools' *overall* pupil/teacher ratios or *general* levels of expenditure on teaching resources.

There was, though, some evidence of a relationship between pupils' science performance and the adequacy of the laboratory accommodation and the technician support in their schools, and also evidence that the higher the schools' levels of expenditure on science education the higher their pupils' science performances.

It was not possible to comment on the possible association between pupils' science performances and teacher qualifications, as there was little variation from region to region or between one type of school and another in terms of the high proportions of science teachers holding a qualification in science (the specific subject area of their individual qualifications had not been requested).

It should be said at this point that where evidence has been shown of an association between pupil performance and any particular school characteristic, this should not be taken to imply a *causal* relationship between the two. Indeed, there is a great deal of overlap between the various links found. The region/location/social class overlap within England has already been mentioned. If we consider levels of science provision in the schools, we again find some overlap between these results and this possible social class influence, given that the survey schools in the South of England in general had higher levels of resource provision and expenditure on science than the schools in other regions.

Significant sex differences in pupils' science performances were found. While there was no difference in the levels of performance of boys and girls for observation skills and for design skills (control of variables, etc), there were *significant* differences in favour of boys for data interpretation skills, in their use of apparatus and measuring instruments and, in particular, for the sub-category requiring the application of concepts of physics. The physics discrepancy has been found many times before, has been shown to be already in existence at a very early age, and is thought to be exacerbated by the tendency of girls to avoid physics when optional course choices are made.

The association between pupils' science performances and the science subjects being studied by them is a difficult one on which to comment at this stage. This is because of the influence on performance of three interacting variables: sex, ability and courses being taken. It can be said that there appeared to be an association between levels of performance and number of science subjects being studied, and that those pupils taking physics and chemistry (with or without biology) were usually the highest achievers. Those pupils taking at least one science subject generally achieved higher performance levels than those taking either a general science course or no science at all.

It is hoped that when further survey data becomes available it may be possible to interpret the curriculum/sex/ability influence on performance in more detail.

10 Summary and discussion of the survey findings

10.1 Levels of performance

The findings reported in this survey give an indication of the levels of performance on a range of science related activities for all pupils in the last year of compulsory schooling, irrespective of the science courses they have been studying. In this final chapter we first summarise the main features of the performance levels in the range of activities tested, then discuss the associations between levels of performance and other variables. Finally we raise for discussion some general questions concerning the educational implications of the findings and plans for future surveys.

Data transformation skills

In tasks related to data handling about 90% of pupils were able to ''read off'' information from graphs, charts and tables with reasonable accuracy, and about 60% were able to perform the basic component skills involved in constructing graphs and charts (such as setting up scales and plotting points). A considerably lower proportion of pupils performed these tasks to a high degree of accuracy.

In questions where pupils had to describe patterns or relationships in data, about two-thirds of the pupils correctly described the relationship when the information was in the form of diagrams or drawings of objects or events. Where the information was presented in a more symbolic form, for example as sets of numerical data in tables or graphs, about half the pupils made correct predictions based on the data, and about a quarter of the pupils gave a correct description of the relationship. (As we know that most pupils were able to read information from graphs, tables and charts with reasonable accuracy it does not appear that it is failure at this that impedes pupils' performance on data interpretation.)

Practical skills

Pupils' use of a number of measuring instruments indicated that levels of performance were affected by the type and complexity of the scale on the instrument. Where the scale was linear and the reading to be taken on a main division, then between 60% and 80% of pupils made a correct reading. When the scale was more complex, either non-linear or requiring interpolation, only about a third of the pupils made a correct reading. Simple measurement techniques were performed adequately by the majority of pupils: however, where more complex techniques or sequences of activity were required the level of performance fell.

The assessment of pupils' ability to estimate physical quantities indicated that about three-quarters of the pupils gave reasonable estimates for quantities of length and time. Fewer than half the pupils gave reasonable estimates for other quantities such as area, force, volume and temperature.

In observation tasks where a number of observations were called for, over three-quarters of the pupils made at least half the observations and about half the pupils made at least two-thirds of the observations that were required or possible. In this type of question there was a tendency for pupils to give answers based on prior knowledge rather than on observations made in the test situation.

Concept application and design skills

The levels of performance on tests of application of science concepts were lower than those for data interpretation even though the questions asked were similar in structure. Overall the levels of performance on applying physics and chemistry concepts were lower than those for biological concepts perhaps reflecting the fact that biological concepts are more familiar and accessible to most pupils. The difference may also be a consequence of the fact that a larger proportion of 15 year olds study biology than study either physics or chemistry. As the results in Chapter 7 indicate, questions requiring the application of more formal scientific concepts involving the application of an exact relationship or the use of a theoretical model were answered correctly by only a minority of pupils.

The levels of performance on questions assessing skills of experimental design appear to be affected by the amount of science knowledge required in the question. In questions which draw only on everyday knowledge, about two-thirds of the pupils correctly responded to tasks which involved such factors as the control of variables or the use of experimental procedures. Where these design questions were set in a scientific context and required pupils to recall and apply their scientific knowledge fewer than half the pupils responded correctly. In longer questions where pupils had to describe how they would conduct a novel experiment about half the pupils described a workable method. A considerably smaller proportion spontaneously incorporated considerations of factors such as the control of variables or sources of error into their accounts.

10.2

Science curriculum groups and science performance

The relationship between pupils' science curriculum experience and levels of science performance at age 15 is clearly a matter of some interest. It is an obvious question to ask how the levels of performance in science are related to the amount and type of science studied. This is a complex question to answer since (as was outlined in earlier chapters) the pattern for science courses followed at this age tends to be different for boys and girls and for pupils of different levels of academic attainment, and this needs to be taken into account before comparisons in levels of performance are made.

Considering first the amount of science studied at this age, the survey indicates that nearly 40% of pupils studied two or more science subjects, half the pupils in the sample studied one science subject and about 10% had discontinued their study of science. The picture is markedly different for girls and boys. Half the boys take two or more science subjects compared with only a quarter of the girls. The pattern of the subjects studied also differs with over 50% of boys studying physics compared with about 20% of girls, and twice as many girls as boys studying biology. Clearly these differences need to be borne in mind when differences in performance between boys and girls and between pupils in different science curriculum groups are considered. In addition, the number of science subjects taken was different for pupils of different abilities. As figure 2.14 indicates, pupils taking 6 or more O-levels tend to take a larger number of science courses; 26% study all three sciences and only 4% take no science at all at this age. By comparison, pupils who are entered for only CSE examinations take fewer science courses, in most cases such pupils are studying only one science course, only 1% of these pupils study all three sciences and 12% take no science at all.

Figure 9.9 in the previous chapter illustrates the pattern of the results for pupils taking different numbers and combinations of science subjects. For all pupils irrespective of their examination group, the more science subjects studied the higher were the levels of performance. This pattern was consistent for all the types of test used in this survey, suggesting that all the abilities assessed are enhanced by additional teaching in science. The one type of ability which was least influenced by science teaching was that of reading off information from graphs and charts, which is a skill that is widely used and hence possibly developed in a range of school subjects including mathematics, geography and social sciences (Holmes et al. 1981). There was, in general, a difference of about 20% in levels of performance between those pupils taking three separate sciences and those studying no science at all. This general pattern was the same for the pupils in the different examination entry groups. Of course, it must be recognised that pupils who have chosen to study two or more sciences are a self-selected sample, possibly with a higher level of attainment and motivation on entry to the science courses than others. However, the findings do suggest that additional time spent on science in the last two years of schooling may have a significant influence on general levels of scientific performance for all pupils. It is surely of interest to note that the difference in levels of performance between pupils studying three sciences and those taking no science at all is of the same order (about 20%) as the difference between pupils in examination entry groups 1 and 4. This result does suggest that with additional teaching in science pupils of lower general levels of attainment can perform as well as those with higher levels of attainment in school who take less science.

The relationships between different science subjects and levels of performance have been reviewed in Chapter 9. In general those pupils studying physics, particularly in combination with chemistry, perform at a higher level than pupils following other course combinations on most types of test. Within the high ability group the pupils taking a general science course performed at a similar level to the pupils taking a single specialist science course except on tests of application in the physical sciences. Analysis of the performance levels within the examination entry group 4 indicated that pupils following a single subject course performed at a consistently higher level than pupils following a general science course. It should be noted, however, that this finding may be an artefact of the broad ability groupings used in this analysis. The difference between the pupils taking a single science and those taking a general science course may not be attributable to differences in the courses followed but may rather reflect differences in the ability and motivation of pupils in those courses.

It was not possible to compare levels of performance of pupils taking an Integrated Science course such as SCISP (which is equivalent in teaching time to two science subjects) with other curriculum groups since the number of pupils in the sample following such a course was too small.

There may be changes over time in the course choices of girls and boys, and of pupils of different abilities, and these will undoubtedly have an effect on the performance levels of the various curriculum groups. It cannot be assumed that the ability indicator used in this survey will continue to be useful, given the possibility of plans to introduce a common examination system at 16-plus. In order to have available some additional information about the general academic abilities of the pupils taking particular science course combinations, the APU mathematics and language monitoring teams are collecting details of the science courses being taken by their survey sample pupils so that the mathematics and language performance levels of the different science curriculum groups can be monitored alongside their science performances.

Analysis of the results for the 1980 surveys (DES 1982b, 1982c) indicated that the highest mean scores in tests of mathematics and language were obtained by pupils studying all three sciences and the lower mean scores tended to be obtained by pupils studying human biology and general science. Separate mean scores for boys and girls in mathematics and language were obtained for the range of science curriculum groups. These indicated that the girls' results in mathematics were either the same or higher than those for the boys for most of the science subject combinations. These findings are consistent with the pattern of 'ability' for pupils in different science curriculum groups reported in Chapter 2.

10.3 Sex differences in performance

The differences in performance between boys and girls in science is a matter of current concern in education. The results reported in this survey may add further detail to a familiar picture. The performance of boys was higher than that of girls in four of the ten types of test: reading off information, using apparatus and measuring equipment, data interpretation and applying physics concepts. Bearing in mind that in general boys have spent considerably more time on science than girls by age 15 it is interesting that differences in levels of performance were not apparent in this survey in other areas such as observation skills or experimental design.

The difference in performance between boys and girls in applying physics concepts has previously been noted for 11 and 13 year old pupils (DES 1981a, 1982a). Particularly given that nearly three times as many boys as girls study physics at this age it is perhaps not surprising to find such a large difference in performance between the sexes on this type of test at age 15. The tendency for boys to choose, or be directed, to study physics and for girls to avoid the subject was illustrated in Chapter 2 when the science subject choices of pupils in the survey were considered, and this difference must further exacerbate the discrepancy in performance which has been shown to be already in existence before pupils make their optional science course choices. It is also possible that this same curriculum difference could also account for the significant sex differences in performance in favour of boys on data transformation and interpretation skills and in using apparatus and measuring equipment, skills which may be developed in the context of a physical science course.

10.4 Performance in relation to school characteristics

The information collected through the school questionnaire gives a general picture of the level of provision of resources for science teaching in secondary schools. It has also enabled some links between provision for science and levels of performance to be explored. These are reported in full in Chapter 9.

There was no evidence of any association between pupil performance and general school characteristics such as the size or age range of the school. However, factors concerned with the level of provision for science teaching, such as expenditure on science, the level of laboratory provision and technician support appear to be related to pupil performance in science; with higher levels of performance being associated with better levels of provision for science teaching.

The findings of the questionnaire indicated that about two-thirds of the survey schools in England had sufficient laboratory accommodation to give all pupils up to age 16 a laboratory-based science course which would meet the HMI criterion of adequacy. A smaller proportion of schools in Wales and Northern Ireland were adequately provided in this respect. The proportion of the survey schools which had both adequate laboratory accommodation *and* sufficient technician support for the laboratories was about 40%.

Estimates given for expenditure on science indicated that schools most often spend up to £2 per pupil per year on materials for science teaching. In fewer than a third of the schools was the sum greater than £3 per head per year. In general about two-thirds of this sum is spent on equipment for practical work; the rest is spent on books and stationery.

About 95% of teachers teaching science in the survey schools held a qualification in science. The proportion was lower for 11–16 comprehensive schools and modern schools whose overall level of provision for teaching science was poorer than for schools of other types. No information was collected in this survey about the specialism of the science subject teachers: this is a factor which will be explored in the future.

The pattern of performance across the regions was generally consistent with the highest performance in the South and Midlands and the lowest in the North, Wales and Northern Ireland. This could reflect the varied contribution of resources for science teaching with the level of laboratory provision and technician support being highest in the South of England and lowest in Wales and Northern Ireland. The lower performances of pupils in Wales and Northern Ireland could also be associated with the proportion of pupils choosing to discontinue their study of science. In England about 8% of pupils in the sample were not following any science course: in Wales the proportion was 11% and in Northern Ireland 26%.

Information on the size of teaching groups for science was not collected in this survey. At age 15 the size of teaching groups for science can vary considerably within a school and for pupils who study more than one science subject it can vary from one subject to another.

The information collected about the schools and the levels of provision for science teaching was restricted in this survey to factors for which objective data were available. It is recognised that other factors such as teaching style, the amount and type of practical work and the general ethos of a science department in a school may have a considerable effect on pupils' performance in science. However, these are more subtle factors to assess. The collection of the necessary information would have been time consuming and would have imposed an additional burden on each school participating in the survey. Such inquiries are possibly more appropriate to in-depth research studies.

It is also recognised that in making an adequate appraisal of the school based factors affecting levels of performance, other influential factors such as the amount of support a pupil receives from home would need to be taken into account. Failure to do this may result in various school effects being masked or misinterpreted.

10.5 Monitoring in science: future plans

This survey is part of a longer term monitoring programme, this being the first of a series of reports of scientific performance at age 15.

In this survey, questions were selected for use in the tests from a larger bank of questions. Although a total of about 400 questions were used this year there were relatively few questions on any particular skill or type of activity. As more surveys are conducted, information from additional questions in each area will be collected. This should enable reports to be prepared which give a more detailed account of pupils' performances in specific areas including analysis of the types of errors made and the difficulties commonly encountered by pupils. It is hoped that reports of this kind will be of use in informing the teaching of specific topics or skills in science.

The link between science performance and the different science curricula followed has been difficult to assess adequately in a single survey, because of the complex interplay of a number of factors. In the future it is planned to explore the issue further by analysing data from more than one year's survey.

In addition, data from further surveys will make it possible to investigate more fully the complex relationships between performance levels and the various school factors affecting pupils' learning experiences in science.

Feedback from the initial surveys will also yield information about the test instruments themselves. This may result in some modifications in the design of the monitoring framework in the future.

10.6 Implications for science teaching: matters for discussion

The performance levels reported for the different sub-categories each assess a distinct type of activity using criteria appropriate to that activity. It is not meaningful, therefore, to make direct comparisons between sub-category scores. It would, for example, be inappropriate to compare a mean performance level of 40% on the use of apparatus and measuring instruments with a mean performance level of over 60% on experimental design and say that pupils are, therefore, better at experimental design than they are at the use of apparatus. Such levels of performance have to be judged in terms of the types of questions and associated mark schemes used to make the assessments. It has been seen as important to include a relatively large number of questions in this report in order to give readers with different interests in the results the opportunity to make their own judgements on the performance levels reported.

Some comments concerning the balance of activities in science lessons can perhaps be made together with suggestions of where more emphasis might be placed.

In general pupils are able to read information correctly from graphs, charts and tables. However, they are less proficient at putting data into these forms. The results of the survey also indicate that more attention might be given to levels of performance in basic practical skills involving more complex measurement techniques and skills of estimation; it should not be assumed that these will be acquired 'in passing' in a science course.

The small proportion of pupils who gave adequate written accounts of an experimental design has been noted. Opportunities for pupils to design experiments, or comment critically on the design of others, are not often given in science lessons. In their recent survey of secondary schools (DES 1979), HMI reported that in only about half the schools visited had pupils been given the opportunity to devise their own experiments. A similar finding is noted in the report of an observation-based study of science classrooms; the researchers stated "over a quarter of the teachers never encouraged the designing of experiments....and....few directions were given to hypothesise or design experiments" (Eggleston et al 1976, p.121). If skills of experimental design are seen to be important, then opportunities may need to be given to develop these skills explicitly. Since experimental design is dependent on knowledge of experimental procedures and on understanding of the scientific concepts relevant to the inquiry, care would have to be taken to select investigations appropriate to the pupils' knowledge and skills. The secondary survey also reports that "in many schools, the acquisition of knowledge was the main feature of the science courses; only rarely was the emphasis on teaching the process of science rather than the subject matter" (DES 1979, p.188).

Despite this reported concentration on the acquisition of knowledge in science lessons the results of this survey indicate that many pupils including the most able have difficulty in *applying* even some simple scientific ideas to everyday situations. It is possible that if the content to be covered in science syllabuses were reduced some of the time available to science teaching could be used to give pupils more opportunities to apply their knowledge and so achieve some mastery over it.

It is likely that low levels of performance in some areas are due to lack of exposure of pupils to the types of tasks involved; additional attention in these areas may significantly affect levels of performance. In other areas, however, low levels of performance may be a reflection not of lack of emphasis in teaching but of the difficulty or complexity of the types of tasks themselves.

This survey can give little guidance on this issue, nor can it indicate which teaching approaches are most effective in enhancing certain skills. To answer such questions would involve detailed studies of teaching in action.

Earlier in this chapter the general levels of performance of 15 year old pupils on science based activities were described. There are two distinct questions that can be asked about these reported levels of performance. First, do they reflect what average 15 year old pupils are capable of if given appropriate teaching? This is not a question that can be answered from the survey data alone; it would require information from classrooms about teaching approaches used, and judgements from practicing teachers. Secondly, do the results reflect levels of performance expected of school leavers by society? Here again judgements are required by various groups of people including parents, educators and employers.

We hope that this report will stimulate discussion of these matters by various groups of people and that together with the reports relating to science performance at other ages, it will contribute to an informed discussion and appraisal of science education at the secondary level.

References

CIPFA (1982) *Education statistics 1980–81 Actuals,* London: The Chartered Institute of Public Finance and Accountancy.

Cochran, W. G. (1977) *Sampling techniques.* New York: Wiley.

Comber, L. C. and Keeves, J. P. (1973) *Science education in nineteen countries: an empirical study.* New York: Wiley.

Davie, R., Butler, N. and Goldstein, H. (1972) *From birth to seven.* London: Longman.

Deming, W. E. (1960) *Sample design in business research.* New York: Wiley.

DES (1978a) "Assessing the performance of pupils." *Report on Education 93.*

DES (1978b) *Science progress report.* Assessment of Performance Unit.

DES (1979) *Aspects of secondary education in England.* A survey by HM Inspectors of schools. London: HMSO.

DES (1980a) *Girls and science.* HMI series: Matters for Discussion 13.

DES (1980b) *Mathematical development. Secondary survey report No. 1.* Report to the DES on the 1978 survey of 15 years olds. London: HMSO.

DES (1981a) *Science in schools: age 11 report No. 1.* Report to the DES, DENI and the Welsh Office on the 1980 survey. London: HMSO.

DES (1981b) *Mathematical development. Primary survey report No. 2.* Report to the DES, DENI and the Welsh Office on the 1979 survey of 11 year olds. London: HMSO.

DES (1982a) *Science in schools: age 13 report No. 1.* Report to the DES on the 1980 survey. London: HMSO.

DES (1982b) *Mathematical development. Secondary survey report No. 3.* Report to the DES, DENI and the Welsh Office on the 1980 survey of 15 year olds. London: HMSO.

DES (1982c) *Language performance in schools. Secondary survey report No. 2.* Report to the DES, DENI and the Welsh Office on the 1980 survey of 15 year olds. London: HMSO.

Eggleston, J. F., Galton, M. J. and Jones, M. E. (1976). *Processes and products of science teaching.* London: MacMillan Education for the Schools Council.

Fogelman, K. (1981). Assessing examination attainment in selective and non-selective secondary schools. In *Publishing school examination results: a discussion..* Bedford Way Papers 5. London: University of London Institute of Education.

Gray, J. (1981). "Are examination results a suitable measure of school performance?" In *Publishing school examination results: a discussion*. Bedford Way Papers 5. London: University of London Institute of Education.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953) *Sample survey methods and theory*. New York: Wiley.

Hobbs, E. D., Boldt, W. B., Erickson, G. L., Quelch, T. P. and Sieben, G. A. (1978) *British Columbia Science Assessment*. Summary report to British Columbia Ministry of Education.

Holmes, P., Kapadia, R., Rubra, G. N., (1981) (Ed. D. Turner), *Statistics in schools 11–16*, Schools Council Working Paper 69, Methuen Educational.

Johnson, S. and Maher, B. (1982) "Monitoring science performance using a computerised question banking system." *British Journal of Educational Technology, 13*, 97–106.

Kellington, S. H., Mitchell, A. C. and Gillespie, A., (1980). "How well can your pupils measure?" *School Science Review, 81* pages 770–774.

Kish, L. (1965) *Survey sampling*. New York: Wiley.

NAEP (1975) *Selected results from the National Assessment of Science: scientific principles and procedures*. Washington: Education Commission of the States.

NAEP (1978a) *Science achievement in the schools*. Washington: Education Commission of the States.

NAEP (1978b) *Three national assessments of science: changes in achievement, 1969–77*. Washington: Education Commission of the States.

Nie, N. H., Hadlai Hull, C., Jenkins, J. G., Steinbrenner, K. and Bent, D. H. (1975) *Statistical package for the social sciences*. New York: McGraw Hill.

Rutter, M., Maughan, B., Mortimer, P. and Ouston, J. (1979). *Fifteen thousand hours*. London: Open Books.

Swan, M. and Jones, O. (1980). "Comparison of students' percepts of distance, weight, height, area and temperature." *Science Education 64(3)*, pages 297–307.

WOED (1982) *The effects on the education service in Wales of recent local authority expenditure policies: an assessment by HM Inspectorate*, Cardiff: HMSO.

Appendix 1

The assessment framework

Introduction

The rationale for the assessment framework was outlined in Chapter 1 of this report, where the main categories and sub-categories were described.

In order to interpret this framework and to develop questions for use in the surveys more detailed documents have been prepared by the team in collaboration with the Steering Group.

A general description has been prepared for each sub-category within the assessment framework. It is this general description that is used by validators in allocating a question to a sub-category (see Appendix 2).

In preparing questions for the bank the team developed a set of "question descriptors" specifying the range of question types to be included in each sub-category pool.

In addition to specifying the types of activity to be tested within each pool, it was also necessary to specify the domain of knowledge within which it was appropriate to set questions. The "List of science concepts and knowledge" was developed for this purpose in collaboration with science teachers. A "List of apparatus and measuring instruments" was also compiled in a similar way.

These documents are included in this appendix. In the first section we give the general description of each category and sub-category as used by validators together with the associated question descriptors which are currently used at age 15. These are followed by the "List of science concepts and knowledge", and the "List of apparatus and measuring instruments" relating to the age 15 tests.

Description of categories, sub-categories and component types of questions

Category 1: Using symbolic representations

In this category pupils are required to use conventional scientific representations. Common mathematical representations will be included.

α *Reading information from graphs, tables and charts*

Pupils are required to read off information from graphs, tables, paragraphs or charts or to label axes.

β *Expressing information as graphs, tables and charts*

Pupils are required to put information presented in words or numbers into various graphical, tabular or symbolic forms, either by constructing such forms entirely or by adding to partially completed ones.

γ *Using scientific symbols and conventions*

Pupils are required to use knowledge of scientific symbols and formulae or of conventions in drawing and labelling.

Given	Demand
Sub-category 1α	
1. A table of a) figures, b) representative symbols, c) non-representative symbols d) a short paragraph	Read off information as directed
2. A schematic representation of a process or series of linked events or relationships	Read off information at directed
3. Vertical or horizontal bar chart	Read off information as directed
4. Pie chart	Read off information at directed
5. Graph	Read off information as directed
6. Graph and statement of what it represents	Read off information as directed
7. a) line graph, b) points located by co-ordinates	Read co-ordinates of a designated point

Sub-category 1β

1. Data and a partially completed table	Add further data to complete table
2. A partially complete schematic representation of a process, series of linked events or relationships and data to add	Add given data to complete schematic representation
3. Data and a partially completed bar chart	Add further data to complete chart
4. Data and a partially completed pie chart	Add further data to complete chart
5. Data and a partially completed line graph	Add further data to complete graph
6. Data in the form of pairs of related quantities	Draw axes, select scale and construct a bar chart
7. Data in the form of continuously varying quantities or pairs of related quantities	Draw axes, select scale and plot points and draw line graph
8. A pair of labelled axes and a) data, b) points to plot, c) data	Construct a) bar chart representing data, b) points, c) line graph — as directed
9. Data represented in four-five different graphical representations	Select the most appropriate representation

Sub-category 1γ

1. Section diagram of assorted objects and a list of names with some redundancy	Identify the objects by matching names to objects
2. A 3-D drawing of an experimental set-up using general laboratory apparatus	Make a conventional section drawing
3. A conventional section drawing of a set-up using general laboratory apparatus	Propose names for the components of the set-up
4. A 3-D drawing of a circuit	Draw a conventional circuit diagram
5. A conventional circuit diagram	Propose names for the components of the circuit.

Category 2: Using apparatus and measuring instruments

In this category pupils are required to use a selection of measuring and other laboratory instruments and apparatus, to estimate and to follow instructions in familiar and unfamiliar situations.

α *Using measuring instruments*

Pupils are required to use units, scales, measuring and other laboratory instruments correctly.

β *Estimating quantities*

Pupils are required to estimate the magnitude of physical quantities by judging their value, or by attempting to take a specified amount of material.

γ *Following instructions for practical work*

Pupils are required to follow written instructions and/or diagrams. Following the instructions may require simple observation, recall of procedures or exact compliance with a set of directions.

Given	Demand
Sub-category 2α	
1. Physical quantity measured by instrument already set up	Give a value with appropriate units
2. Physical quantity, object, or event, and instrument(s) for measuring or observing it	Employ appropriate instrument(s) to give an answer or give a value with units or leave a measured quantity
Sub-category 2β	
1. Objects or events and units of physical quantities	Give a value
2. Supply of material and an amount specified	Leave or indicate the right amount of material

Sub-category 2γ

1. Instructions, in the form of a conventional diagram, for setting up apparatus for a stated purpose	Set up apparatus
2. Detailed instructions for completing an unfamiliar task	Comply exactly with instructions in order to complete the task
3. Detailed instructions for completing an unfamiliar task	Comply exactly with instructions in order to leave a product
4. Instructions for task that explicitly refer to standard techniques used in laboratories	Follow instructions, using the correct procedures, in order to complete the task
5. Instructions for task that require standard techniques used in laboratories. Reference to these is implicit	Follow instructions, recalling and using the correct procedures, in order to complete the task

Category 3: Using observations

In this category pupils have to decide for themselves what are relevant observations and then record or make use of them.

α *Using an identification key*

Pupils are required to identify objects using a branching key.

β *Observing similarities and differences*

Pupils are required to make comparisons either between two or more objects or events, or between successive stages of an event. They may have to decide what is important and relevant in the comparisons and record their decisions in a variety of ways.

γ *Interpreting observations*

Pupils are required to collect information by observation, often by making comparisons or noting changes, and then use what they take to be relevant to go beyond their observations to find patterns, explain what they observe, or predict.

Given	Demand
Sub-category 3α	
1. Objects (or photographs of objects)	Identify objects
Sub-category 3β	
1. Objects (or photographs or drawings of objects)	Identify rules used to classify the objects and/or sort objects into an alternative set of groups
2. At least two objects (or photographs)	State as many differences as possible
3. At least two objects, photographs or events	State a specified number of similarities and differences
4. Event	Make a record of change
5. Objects (or photographs of objects) and drawings	Select the matching drawing
6. Object	Make a scientific drawing
7. Object (or photograph) and another photograph which is part of the object (or first photograph)	Identify where the part comes from
8. Set of photographs or drawings forming a sequence	Arrange the remainder in the correct order

Sub-category 3γ

1. Event	Make a record of observations and create a hypothesis
2. Event and a list of explanations	Make a record of changes and select the appropriate explantion(s)
3. Event	Make a record of changes and make a prediction consistent with the data
4. a) At least two events, objects, photographs or drawings	a) Note the difference and make a prediction consistent with the data
b) At least two events, objects, photographs or drawings and a list of predictions	b) Note the difference and select a prediction consistent with the data

Category 4: Interpretation and application

In this category pupils are required to interpret or apply information presented in a diagrammatic, tabular or written form. The questions are of two main types: those in which all the information necessary for an answer is in the question itself, and those which require the recall and use of science concepts and knowledge.

* * * * *

Questions in which the information or concepts needed to answer the question are either assumed to be within the everyday experience of the pupil, or given in the question itself.

α *Describing and using patterns in information*

Pupils are required to describe, select or use patterns derived from information given or to use generalisations stated in the question. This information may be presented in pictorial, graphical, tabular or written form.

β *Judging the applicability of a given generalisation*

Pupils are required to assess the extent to which a suggested generalisation or hypothesis applies to a given situation(s).

γ *Distinguishing degrees of inference*

Pupils are required to select a statement which makes the fewest additional assumptions about a verbal or pictorial account of an event.

Given	Demand
Sub-category 4α	
1. Data (to be sufficient to answer question without prior knowledge) in pictorial, tabular or graphical form	Describe pattern based on data alone (could involve pupils in drawing sketch graphs to determine pattern)
2. Data (to be sufficient to answer question without prior knowledge) in pictorial, tabular or graphical form	a) Describe pattern and use to make prediction, or b) make prediction giving reasons based on pattern
3. Data (to be sufficient to answer question without prior knowledge) and a minimum of four predictions	a) Describe pattern and use to select prediction or b) select prediction and give reason for choice
4. Data (to be sufficient to answer question without prior knowledge)	Make predictions/explanations based on data alone
5. Data and a minimum of four predictions	Select prediction consistent with data
6. Data sufficient to make required prediction plus a minimum of four predicitions	Select pattern and use to select prediction consistent with data
7. Data and a minimum of four predictions/explanations	Select all the predictions/explanations which are consistent with the data.
Sub-category 4β	
1. Data and hypothesis/explanation	Assess validity of presented hypothesis/explanation in relation to data
Sub-category 4γ	
1. A "snapshot" of an event (in words or line drawing) and five possible accounts of the event	Select the one which demands the fewest additional assumptions

Category 4: Interpretation and application (*continued*)

Questions which require the recall and use of science concepts and knowledge (as specified in the List of Science Concepts and Knowledge).

δ *Applying science concepts to given situations*

Pupils are required to apply science concepts to: make or select predictions or deductions; give or select explanations; Assess explanations or hypotheses in the situation given.

ε *Generating alternative hypotheses*

Pupils are required to generate *alternative* hypotheses each of which must be consistent with information given and with science concepts.

Given	Demand
Sub-category 4δ	
1. Data	a) Describe pattern based on data *and* accepted concepts and use it to make predictions — or b) make predictions giving reasons
2. Data	Make predictions based on data and accepted concepts
3. Data and a minimum of four predicitons	Select prediction and give reason based on data and accepted concepts
4. Data and a minimum of four predictions	Select predictions based on data and accepted concepts
5. Data and a description of an event or situation	Give an explanation consistent with the data and accepted concepts
6. Data, description of an event or situation; and hypothesis/explanation	Assess validity of hypotheses/explanation in relation to data and accepted concepts
7. Data, description of event or situation and a minimum of four hypotheses/explanations	Select the best hypotheses/explanation in relation to data *and* accepted concepts
8. Data, description of event or situation and a minimum of four hypotheses/explanations	Select all the hypotheses/explanations which are consistent with the data *and* accepted concepts
Sub-category 4ε	
1. Data, description of event or situation where there is no single obvious explanation	Generate alternative hypotheses/explanations consistent with data *and* accepted concepts.

Category 5: Design of investigations

In this category pupils are asked to formulate proposals for experimental investigations and to appraise or construct plans for conducting such investigations

α Assessing testable statements

Pupils are required to compose or select a proposition in a form that can be tested by systematic observation, given either a general statement or several alternative statements or propositions.

β Assessing experimental procedures

Given a testable proposition, pupils may be required to select from or criticise or explain or complete a set of procedures and/or variables for testing it. Pupils may also be asked to select the testable proposition that a given procedure will test.

γ Devising and describing investigations

Given a proposition in a testable form, pupils are required to devise and describe a procedure to test it, including the identification and selection of relevant variables and the selection of equipment.

Given	Demand
Sub-category 5α	
1. Four or five statements, one of which:	
a) can be tested scientifically	Select the statement which can be tested scientifically
b) cannot be tested scientifically	Select the one which cannot be tested scientifically
2. A general statement or opinion	Rewrite as a (number of) statement(s) each of which can be tested
Sub-category 5β	
1. A proposition in a testable form and at least four specified variables (or procedures relating to specified variables)	Select which variables must be varied or controlled (or which procedures must be followed) to allow for testing
2. A proposition in a testable form and no specified variables	State variables to be controlled to allow testing
3. A proposition in a testable form and a procedure to test it	State the functions of the component parts of the procedure
4. A proposition in a testable form and at least four component procedures	Propose the best sequence
5. A proposition in a testable form and a procedure to test it	Criticise the procedure
6. A proposition in a testable form and a procedure to test it	Select the functions of the component parts of the procedure
7. A procedure and at least four statements about what it might test	Select the statement(s) which the procedure would test
Sub-category 5γ	
1. A proposition in a testable form	*Plan* a procedure to test, resolve or determine the point at issue (taking into account control of variables and selection of equipment)
2. Two propositions fitting the same data	*Plan* a procedure to distinguish between the two propositions

List of concepts and knowledge

Concept Areas

A. *Interaction of living things with their environment*

1. Interdependence of living things
2. The physical and chemical environment
3. Classification of living things
4. Physical and chemical principles needed to interpret life phenomena

B. *Living things and their life processes*

1. The cell
2. Nutrition
3. Respiration
4. Reproduction
5. Sensitivity and movement

C. *Force and Field*

1. Movement and deformation
2. Properties of matter
3. Forces at a distance
4. The Earth in space

D. *Transfer of energy*

1. Work and energy
2. Current electricity
3. ''Waves''

E. *The classification and structure of matter*

1. States of matter
2. Pure substance
3. Metals and non-metals
4. Acids and bases
5. Periodic table
6. Atomic model

F. *Chemical interactions*

1. Solutions
2. Reactivity
3. Properties of a chemical reaction
4. Some chemical reactions

A INTERACTION OF LIVING THINGS WITH THEIR ENVIRONMENT

A1 Interdependence of living things

Virtually all organisms are dependent on the presence and activities of other organisms for their survival.

Green plants use energy from the sun to make food by "photosynthesis". During this process the oxygen that is used by animals and plants is produced.

Some animals eat plants and some eat other animals, but all animals ultimately depend on green plants for food. This relationship can be illustrated by a "food web".

Any alteration in one part of the "food web" may affect many other parts of the "food web".

An alteration in the "food web" can be as a result of a change of balance between "consumers" and "producers", or of a change in the inorganic environment.

Organisms live in "communities" in which each organism has a place in which it is best adapted to survive.

Competition and predation tend to maintain the balance of populations within a "community".

A2 The physical and chemical environment

The physical and chemical conditions necessary to life occur on or near the earth's surface.

The composition of air tends to remain relatively stable as a result of cyclic processes involving living things.

Water is essential both as a raw material and as the "solvent" involved in transport and chemical processes within living things.

Soil is a mixture that includes rock particles, humus, water, air and living things.

The return of substances to the soil by the death and decay of living things helps to maintain soil fertility.

Changes in the physical environment due to seasonal cycles are often matched by changes or events in the living world, such as fruiting or mating.

A3 Classification of living things

There are many different plants and animals showing a variety of ways of carrying out life processes.

There are many different criteria that can be used to sort living things, eg. structure, behaviour, etc.

Plants are distinguished from animals by a number of criteria including cell structure and method of obtaining food.

Plants are classified into two major groups — flowering and non-flowering.

Animals are classified into two major groups — invertebrates and vertebrates.

There are five main groups of vertebrates: fish, amphibia, reptiles, birds and mammals.

A theory about how organisms have "evolved" is currently used to classify the living kingdom.

A4 Physical and chemical principles needed to interpret the phenomena

Diffusion in living organisms occurs in an aqueous medium. Water particles move by diffusion as do particles of substances dissolved in water.

Osmosis is a special case of diffusion taking place across a "selectively permeable" membrane which usually allows the passage of water particles only.

The area of its surface affects the gain or loss of energy and matter from an organism or cell.

The ratio of surface area to the mass of an organism is critical for its survival and will reflect its relationship with its environment.

Animals that have a stable body temperature have a number of mechanisms for the control of loss of heat energy.

B LIVING THINGS AND THEIR LIFE PROCESSES

B1 The cell

The cell is the basic unit of most living things.

A "typical" cell has a nucleus, cytoplasm and a cell membrane.

The nucleus is the controlling centre of the cell and contains "hereditary" material.

The cell membrane acts as a "selectively permeable" barrier.

Plant cells have a cell wall outside the cell membrane.

Cells show a wide range of adaptation to the particular function they perform (eg. in shape, inclusions, size).

An organism may be formed from one or many cells.

In multicellular organisms cells are usually grouped to form tissues and tissues grouped to form organs or systems.

Normal cell division, "mitosis", results in two identical cells and leads to either growth or asexual reproduction.

B2 Nutrition

Food is used by all organisms as a source of energy and of raw materials for growth and reproduction.

Green plants take in simple substances from their environment and build them into more complex substances.

All food substances must be dissolved before the cells can use them.

Natural foods contain fats, proteins and carbohydrates.

In humans, an adequate diet will contain a sufficient and balanced amount of these foodstuffs as well as water, vitamins, minerals, salts and roughage.

Digestion is the process in which foods taken into the body are broken down by mechanical and chemical means into simple soluble substances for absorption and transport.

Surplus food is stored in an insoluble form as fat or carbohydrate.

Enzymes change the rate of certain chemical reactions under specific conditions.

All organisms are "structurally adapted" to the kind of food they need.

The structure of plants reflects their ability to manufacture organic foods using absorbed light energy.

All living things produce waste materials in carrying out life processes.

B3 Respiration

Most living things take in oxygen from their surroundings to be used in the process of respiration.

All living things respire in order to make use of energy stored in food.

In respiration, energy is transferred from organic food material with the release of carbon dioxide.

In the case of "aerobic" respiration, water is also produced.

Many living things have special organs which enable gas exchange to take place efficiently.

Gas exchange organs have a large surface area for their volume, are damp, and often employ "mass flow" aids.

B4 Reproduction

Living things produce offspring of the same species as themselves.

Asexual reproduction involves only one parent and the offspring are identical to that parent.

There are many different types of asexual reproduction but they all involve "mitotic" cell division.

Sexual reproduction involves two parents and generally results in offspring that differ from both parents and from each other. This leads to variation.

Sexual reproduction involves the production of "gametes", ova and sperm or pollen, produced by a special type of cell division ("meiosis"), and their fusion to form a zygote.

The human sexual cycles and changes are controlled by hormones.

The life cycle of all organisms repeats itself every generation.

B5 Sensitivity and movement

All living organisms have means of receiving information from their environment.

Special organs/cells are concerned with receiving different kinds of "stimuli". These are called sense organs/cells.

In humans the senses include sight, touch, hearing, taste and smell.

The senses developed by an organism show an adaptation to its particular way of life.

The response to a "stimulus" often results in movement or, in higher plants, in directional growth.

Voluntary movement in many animals is brought about by contraction of muscles attached to a skeleton.

C FORCE AND FIELD

C1 Movement and deformation

The average speed of an object is found by dividing the distance moved by the time taken.

When an object travelling in a straight line changes its speed it is said to accelerate. The size of the acceleration is the change in speed divided by the time.

An object which is still or moving with uniform speed in a straight line has no unbalanced forces acting on it.

An unbalanced force makes an object accelerate and the acceleration is greater when the force increases. For a given unbalanced force, the acceleration is smaller for larger masses.

Rotation is produced when an object has a pair of parallel forces which are not in line acting on it; the turning effect, or "moment" of a force about a point is the magnitude of the force multiplied by the distance of the line of action of the force from the point.

Objects are deformed by and can break under the action of "opposed" forces. The force which an object can stand before breaking depends on its shape as well as the material it is made of.

If materials recover their original shape after a force which caused deformation has been removed, the material is called elastic.

C2 Properties of matter

Different substances have different masses for equal volume; the mass of a unit volume of a substance is its density.

If the volume of a given mass increases, its density decreases.

Pressure is the magnitude of the force per unit area acting on a surface.

Pressure at a point in a gas or liquid depends on the depth of the point below the surface and the density of the gas or liquid.

Pressure at a point within a gas or liquid acts equally in all directions if there is no movement in the gas or liquid.

Particles of a substance are in constant motion. Diffusion of a substance is due to the random motion of individual particles (which may be molecules, atoms or ions).

If the pressure of a fixed mass of gas is increased, its volume decreases provided the temperature stays the same.

The pressure exerted by a gas on a surface is the result of continual bombardment of the surface by many particles.

Most substances expand as their temperature increases.

Rigid objects completely immersed in a gas or liquid displace a volume of gas or liquid equal to their own volume.

There is an upward force on an immersed object which is equal to the weight of gas or liquid displaced; a floating object displaces a weight of gas or liquid equal to its own weight.

Surface tension and capillarity may be accounted for in terms of intermolecular forces.

C3 Forces at a distance

Magnets attract and repel other magnets and attract magnetic substances.

The region in which a magnetic effect can be detected is called a magnetic field.

There is a magnetic field surrounding the Earth.

Magnetism is induced in some materials when they are placed in a magnetic field.

An electric current in a conductor produces a magnetic field round it.

There is a force on a current-carrying conductor in a magnetic field so long as the conductor is not parallel to the field.

The force on a current-carrying conductor in a magnetic field increases with the strength of the field and with the current.

Relative motion of a conductor and a magnetic field can be used to produce an electric current.

Electric charges are separated when certain materials are rubbed against one another.

Similarly charged objects repel each other, and oppositely charged objects attract each other. The force between such objects is increased by bringing the objects closer and by increasing the charges on them.

C4 The Earth in space

The weight of an object on the Earth is the force with which it is attracted to the Earth.

The weight of an object may vary from place to place, but its mass does not change.

The weight of an object acts through its centre of gravity. The position of the centre of gravity affects the stability of the object.

The apparent movements of the sun, moon, planets and stars follow a regular pattern.

The Earth spins on its axis; this gives rise to night and day.

The Earth revolves in an orbit around the sun with its axis tilted with respect to this orbit; this accounts for seasonal changes.

The moon orbits the Earth as a natural satellite.

D TRANSFER OF ENERGY

D1 Work and energy

There is a variety of sources of energy such as fuels (including food), other chemicals, deformed springs, capacitors and objects at a height.

Energy can be changed from one form to another but can never be created or destroyed. At each change some energy becomes less available for doing useful work.

Moving objects have energy which is transferred when they are stopped.

An object increases its "potential energy" when it is raised from its original position to a greater height above the Earth's surface.

Work is done (energy is transferred) when a force moves its point of application; work is measured by the product of the force and the distance moved in the direction of the force.

The hotter a substance is, the more energy its particles have.

Different substances conduct heat at different rates. Conduction of heat can be explained in terms of the kinetic theory of matter.

The expansion of material on heating can be explained in terms of the additional energy of motion of its particles.

D2 Current electricity

A complete circuit of conducting material is necessary for a steady current to flow between the terminals of a battery or a d.c. power supply.

Electric current is a flow of charged particles.

Some materials conduct electricity better than others. Bad conductors are known as insulators.

Electric currents add up ''algebraically'' at a junction in a circuit.

D3 Waves

Objects are seen because of the light which they give out or reflect.

The path of light travelling in a uniform medium can be represented by straight lines.

Light is reflected regularly at a mirror-like surface.

Light is ''refracted'' at the boundary between two different media which both transmit light.

''Refraction'' and reflection give rise to the formation of images.

Sound can be heard when objects vibrate in a medium.

Sound requires a medium for its transmission.

In general, the loudness of a sound is increased when the amplitude of vibration of the source is increased.

When the frequency of vibration is increased, the note sounds higher.

Different colours of visible light have different wavelengths. Other types of radiation share some of the properties of light and can be located on an extended scale of wavelengths.

E THE CLASSIFICATION AND STRUCTURE OF MATTER

E1 States of matter

In general a substance can be classified either as a solid or a liquid or a gas.

The behaviour of substances can be explained if it is assumed that matter is made of minute particles.

Solids have a definite shape and volume. They behave as if their particles are held together, in regular arrangements.

Liquids have a definite volume and surface but no fixed shape. They behave as if their particles are closely packed but free to move.

Gases have no definite volume or shape. They behave as if their particles are free to move independently of each other.

A change in state does not involve a change in the chemical composition of the substance.

Changes of state caused by heating can be reversed by cooling and vice-versa.

Changes of state always involve a transfer of energy.

E2 Pure substance

A pure substance may be obtained from a mixture using one of several techniques, including evaporation, distillation, chromatography, filtration, sublimation and crystallisation.

A pure substance is recognised by its characteristic chemical and physical properties (at S.T.P.) eg. m.p., b.p., density and behaviour with other substances.

Different pure substances need different amounts of heat to give the same rise in temperature in a given mass.

Pure substances may be classified into elements and compounds.

An element is made up of only one kind of atom ie. of fixed atomic number.

A compound is formed when two or more different elements are chemically combined.

E3 Metals and non-metals

Materials can be classified into groups in many different ways. One way is by sorting into metals and non-metals.

In general, metals can be distinguished from non-metals by their characteristic physical properties. These include high m.p. and b.p., shiny appearance and conduction of heat and electricity.

Another way of distinguishing metals from non-metals is by their characteristic chemical properties. These may include the nature of the oxide and the formation of positive ions.

Some metals do not exhibit all these characteristic properties.

Some elements show both non-metallic and metallic characteristics.

E4 Acids and bases

Compounds can be classified into groups by their different properties. Acids and bases are two such groups.

Acids and bases have a characteristic effect on the colour of indicators.

Bases which are soluble in water are called alkalis.

In aqueous solution the degree of acidity depends on both the substance and its concentration.

The degree of acidity is expressed on a pH scale from 0 to 14.

A neutral solution has a pH value of 7; acidic solutions have pH values less than 7, and alkaline solutions have pH values above 7.

A neutral solution can be obtained by adding an acid to an alkali until the pH value is 7.

E5 Periodic table

Families of elements with similar chemical behaviour can be identified.

Patterns and trends exist in the behaviour of the elements.

The characteristic properties of the elements show a periodic dependence on atomic number.

E6 Atomic model

The atom is the smallest, characteristic uncharged particle of an element.

The behaviour of atoms can be explained if it is assumed that they are composed of a heavy positive nucleus surrounded by negative electrons.

Atoms are characterised by a symbol and the atomic number (number of protons).

An uncharged bonded group of two or more atoms is called a molecule.

An ion is an atom or bonded group of atoms bearing electric charge(s), positive or negative.

F CHEMICAL INTERACTIONS

F1 Solutions

Some substances dissolve in water; others do not, but may dissolve in other liquids.

A liquid which will dissolve a substance to form a solution is called a solvent.

The substance which dissolves in the liquid is called a solute.

At a given temperature and pressure, the mass of solute which will dissolve in a given volume of solvent is limited and fixed.

For most solids solubility increases with increasing temperature.

A saturated solution of a solute in a solvent is one in which no more of that solute will dissolve, at a given temperature and pressure.

The mass of solute which is dissolved in a given volume of solvent determines the concentration of the solution.

F2 Reactivity

A more reactive element can be used to extract a less reactive element from one of its ores or compounds.

The reactions of metals with water or dilute acid can be used to place them in a reactivity series.

The reaction of one metal to replace another from an aqueous solution of its salts may be predicted from this reactivity series.

The polarity of electrodes in a chemical cell can be deduced from the position in this reactivity series of the elements from which they are made.

F3 Properties of a chemical reaction

A chemical reaction occurs when one or more different substances are formed from one or more original substances.

Most chemical reactions are initiated by an input of energy.

The total mass of reacting substances in any chemical reaction is the same as the total mass of products.

Chemical reactions take place at varying and various rates.

The rate of a chemical reaction may depend on several factors eg. concentration, pressure and temperature.

A catalyst is a substance which changes the rate of a chemical reaction without being used up in the reaction.

F4 Some chemical reactions

On heating some compounds change colour due to loss of water. Often these changes are easily reversed.

When elements react with oxygen only they usually form oxides; this is an example of an oxidation reaction.

When a compound changes by losing oxygen, this is an example of a reduction reaction.

The oxidation of a metal by atmospheric oxygen is an example of corrosion.

Fuels such as coal and oil are formed by the gradual decay of plant remains under high pressure.

Large amounts of energy can be transferred from these fuels when they react with oxygen. This is an example of a combustion reaction.

When an acid reacts with an alkali or a base, a salt and water are formed.

When an acid reacts with some other substances eg. with metals or carbonates, a salt and a gas are formed.

The process of electrolysis involves ("electron transfer") reactions which are used extensively in industry.

List of apparatus used at age 15

Item	*Uses to which item can be put in test situations*
1. Test tubes	holding liquids/powders performing chemical experiments in need to know safe method of heating a test tube.
2. Test tube racks	holding test tubes
3. Stirring rod	use for agitating use for mixing to dissolve
4. Test tube holder	holding test tube in flame
5. Rule (meter, other)	measuring distances selecting appropriate scale using procedure for finding small distances
6. Bungs	to stopper containers to make air-tight (rubber only) rubber glass seal
7. Glass tubing	assembly (e.g. "scrubbing flasks" "delivery tubes")
8. Lever arm balance	find weight/mass correct for zero error read with accuracy
9. Stop clock	timing length of an event(s) number of event(s) in a given interval assessing average time interval
10. Thermometer	assess temperature wait for equilibrium select instrument appropriate for range take a sequence of timed readings
11. Petri dish	holding specimens
12. Bunsen burner	heating correct use of collar
13. Tripod	carry heated object
14. Gauze	appropriate use of heat spreader
15. Filter paper	to filter (correct folding technique) chromatography weighing powders
16. Funnel	to filter to pour accurately
17. Dropper bottle	to add reagent with control use of stopper non-contamination techniques
18. Beaker	hold liquids/solids crude measure of volume heating liquids use as water bath

19. Conical flask	holds liquids mix liquids with liquids (swirl) mix liquids with solids heat liquids
20. Measuring cylinder	to measure liquid volumes to assess irregular solid volumes select appropriate size
21. Teat pipette	to add liquids to transfer liquids
22. Tongs	to handle hot/dangerous liquids
23. Spatula	to add solids non-contamination techniques
24. Hand lens	to magnify to identify
25. Microscope	to magnify to focus to draw from (using open eye technique) to illuminate (understage, top stage)
26. Slides and coverslips	to prepare slides full technique for temporary mounts
27. Forceps (tweezers)	to hold objects either small or dangerous to pick out "dissecting"
28. Lamp (pea bulb)	to test a circuit torch
29. Cell (electric)	to power a circuit
30. Switch (circuit breaker)	for circuitry
31. Wire	for attaching objects to use in a circuit
32. Magnet	to identify magnetic metals to display magnetic field (using iron filings)
33. Compass	to identify direction of magnetic field
34. Manometer	to find a gas pressure
35. Newton meter (spring balance)	as a balance to measure force
36. Clamp stand	used as a general purpose holder
37. Clamp boss	used as a general purpose holder
38. Clamp	used as a general purpose holder
39. Round bottomed flask	to boil liquids in
40. Lab. power pack	to supply variable electrical power
41. Sand tray	to spread heat gently
42. Crucible	for strong heating

43. Pipe clay triangle	to support crucible
44. Burette	to measure volumes of liquids accurately
45. Syringe	transfer of measured quantities of liquid test the type and composition of a gas
46. Ammeter	measure currect in amps and milliamps
47. Voltmeter	measures the potential difference
48. Splint/taper	to light Bunsen burner testing of gases
49. Spring	extension and compression
50. Tile	spotting out of liquids displaying specimens
51. Universal indicator	pH indicator (use with soils etc.)
52. Litmus	pH indicator
53. Limewater	detection of carbon dioxide
54. Bicarbonate indicator	detection of carbon dioxide

Appendix 2

Development and selection of the science questions

The categories and sub-categories of the assessment framework were outlined in Chapter 1 of this report, and were reviewed more fully in Appendix 1. It was with this framework in mind that questions have been developed for use in the monitoring programme, and questions were subjected to a number of procedures before being considered suitable.

Each question was first reviewed (''shredded'') by a small group of people (usually including one or two team members and practising teachers) who commented on its suitability with regard to content, format, language, and so on.

Once considered suitable for presentation to pupils of an appropriate age, questions were grouped together into ''test packages'' and administered to a small number of pupils (about 30) in two or three volunteer schools. On the basis of these pupils' responses mark schemes were developed.

Questions belonging to the same sub-category should ideally share a common mark scale if they are to make equal contributions to the sub-category test score when they are chosen for use in a survey. However, it was decided that, given the variety of question types which could be written for some sub-categories, it would impose artificial constraints on the question developers if they were to be required to create questions with a common maximum mark in mind. The original (''raw'', ''natural'') mark allocation is therefore allowed to vary from question to question so that one question might, for example, carry a maximum mark of 6 while another might carry a maximum of 4 or 5 or 3. This means that if equal weight is to be given to each question in a test the natural scores achieved by individual pupils on individual questions have to be mapped onto a common mark scale. The way in which such transformations are to be carried out is decided by the question developers. In most cases, a common scale is used with marks ranging from 0 – 3. For some sub-categories, (4δ for example) two groups of questions had to be distinguished — those with a natural maximum mark of 1 whose scores are not transformed (usually multiple-choice questions), and those with natural maximum marks of 3 or above whose scores would be transformed when necessary onto a 0 – 3 scale.

The validity of the questions has been judged by a group of volunteer science educators. Any particular question was scrutinised by seven people, working independently, who either assigned it to a specific sub-category or else indicated that such an assignment was not in their opinion possible. A question was considered validated if five out of seven ''validators'' assigned it to the same sub-category as that intended by the team. Questions could be rejected completely at this stage.

Questions were also trialled on a large scale, by administration in test packages to between 100 and 200 pupils for written tests and to about 90 pupils for practical tests. The criterion of suitability for use in a future sub-category test is that the correlation between pupils' transformed scores on an individual question and their total test scores for the relevant sub-category should reach 0.3. Questions failing this statistical criterion would be rejected, as would those questions which proved too complex or expensive to administer on this scale.

Every suitable question has attached to it a series of descriptive "labels". These labels indicate, for example, the relevant sub-category and question descriptor, the question's maximum natural and transformed marks, its transformation algorithm, required apparatus or other resources, and so on. In addition, a synopsis and a number of keywords are provided.

The question bank which contains all the available science questions is in two parts: the question texts, diagrams and mark schemes are manually catalogued, while the descriptive labels associated with the questions are computer-stored. These labels, in combination, describe the totality of questions available. Each question is given a unique identification number and this provides the link between the two sections of the bank.

The computer-storage of question labels serves a major function in the monitoring exercise in allowing the automatic production of random selections of questions. These selections form the sub-category tests to be used in the surveys.

An appropriate set of labels is used to identify all the questions which together form the "pool" for a particular sub-category. From within this pool questions are randomly selected to produce the collection of questions known as the "sub-category test". Such random selection is usually subject to a minimum, but necessary, number of constraints to avoid, for example, the chance production of a question selection which may carry too great a burden in terms of apparatus provision or testing time, or which may contain unsuitable emphasis on one kind of question or task at the expense of some other.

Any individual pupil would not be required to attempt the complete set of questions relating to a single sub-category within one survey, so that the complete sub-category test is in practice sub-divided into two or more shorter tests, and any selected child will take only one of these within his/her total hour-long test package. In 1980, the sub-tests within one sub-category were created separately and independently, and were of similar composition in terms of the number and variety of questions they contained.

Further information about the question banking system and the test construction procedures is to be found in Johnson and Maher (1982).

Appendix 3

The organisation of the group practical tests

The 1980 practical tests

For the 1980 survey 46 administrators were recruited for the testing in England, Wales and Northern Ireland. 34 were used in England and they went to 155 schools, administered 310 circuses, tested 2,790 pupils and travelled a total of 14,200 miles.

Every one of the administrators coped with arduous tasks very well and the team greatly appreciates the time and effort they put into making the practicals successful.

The circus organisation

In the practical assessments at age 15/16, the pupil is presented with a "package" of questions in the form of a circus.

A circus is made up of stations, at each of which the pupils spend 6 minutes. There is usually one question at a station, but there can be up to three shorter questions. The questions are written so that at least 80% of the pupils will have finished within the six minutes allowed.

Each circus is used to test nine pupils at a time. Two circuses are run in each school. Administrators, who take with them all of the equipment and stationery needed, visit the school, and with the help of a teacher from the school, set up and run the circuses. The tests last for about 1 hour so both circuses can be run in a morning session, but a variety of factors including the time it takes to set up the apparatus, complete the marking of some of the tasks, and the dismantling and packing away, can occasionally cause the testing to continue into the afternoon. A laboratory is needed for the tests because any or all three of the services (gas, water and electricity) may be required for a circus.

Selection of testers

When the provisional sample of schools had been drawn Local Education Authorities that were suitably located were approached with a request for the names of two or three appropriately qualified teachers. Due to the combination of skills required of the administrators these were all science teachers and most were either heads of department or deputy heads.

Each administrator was recruited for one week of testing in schools and two days training. Provision was made to pay for a replacement teacher during their absence.

The training programme

The administrators were required to attend a two day training session one week before the survey. Every aspect of the running of the circuses was rehearsed. Particular attention was paid to the setting up servicing and marking of the questions. By the end of the session every administrator was equipped to run their circuses in a reliable and consistent manner.

Running the circus

The circus was run with copies of the question left in a pile at the station. The pupils picked up the question sheet and joined it to their collections of paper with a treasury tag as they passed through the station. This allowed flexibility in station ordering and so allowed for variation in laboratory layout. (It had been found in trial runs that the question order did not affect the performance of pupils.)

The administrators had to follow a pre-defined set of procedures when running the circuses. These are described in the copy of "Instructions to administrators" which follows. This was given to the administrators to act as an aide-memoire, and was used to keep the tests consistent between administrations and administrators.

The administrators were also given detailed instructions about how to set up the apparatus at each station and how to "service" it. Every pupil had got to see the apparatus in the same state as the first pupil, and this entailed rearranging or replacing some items of equipment between each use.

All the apparatus was supplied, except for microscopes, and packed in custom-made cardboard boxes for easy and reliable handling during the test period.

APU Science: instructions for administrators of practical survey 1980

To set up circus:

Arrange stations in the laboratory so that:

a) questions needing services such as gas and water are suitably placed,

b) the route through the circus is simple to follow for the pupils,

c) questions needing servicing or marking are convenient for you.

NB The sequence of the stations is not important. You do not have to keep to the "order" on the circus sheet.

At each station put out:

1) the apparatus — in standard arrangement

2) the pile of questions — normally to left of apparatus

NB If there is more than 1 question at a station, use the prompt card to show this. (Normally the shorter, or easier, question to be done first, is put on the left.)

3) the station number — so that it is clearly seen

4) an instruction page ie. the rubric

5) a treasury tag

6) the control card appropriate for starting at that station

Train the "resident" teacher. Usually the teacher is involved in using a check-list to mark the pupils' performance at one of the stations.

Organising the pupils

The resident teacher should have a list which matches pupil names to APU numbers. You have front pages with numbers only. Get them to give the right numbered front pages to the correct pupils.

Arrange the pupils in numerical order, lowest number first.

Tell them to go to the stations, lowest number to station 1, next lowest to station 2 etc. Not to start yet. (Stations numbered greater than 9 will remain empty at first.)

Starting the circus

Tell the pupils who you are and what you are doing.

Tell them that the tests are confidential but their results are important as they are representative of the age group.

Read out the instruction sheet, ie. the rubric, with explanations. Get pupils to tag up their control card, front page, and instruction page.

Tell them they have 6 minutes at each station, and that they can start work at the new station immediately they have moved.

Say ''begin'' and fill in the time sheet.

Give the pupils 7 minutes for the first two stations, and 6 minutes after that. (This allows time for the pupils to ''settle'' and confers no advantage to the pupils at the stations at the start of the circus.)

Two minutes before the end of the station, give a warning.

At the end of the 6 (7) minutes, (actually at 5′50″), say ''Tag up your questions. Get ready to move to your next station. Move please.'' Do this each time.

If a hiatus occurs see below.

Finishing the circus

Tell the pupils to stop at the end of the final 6 minutes.

Staying where they are ask them to (i) remove the control card; (ii) put their control letter on front page; (iii) check that grids of subjects, sex, date of birth are filled in; (iv) leave the scripts where they are.

Thank them for taking part and dismiss them. Most pupils enjoy the circuses. If they are interested (and you have the time) you may of course talk to them about the questions and the reason for the tests etc. It is the only direct feedback that they will ever get.

Hiatus

If a problem of any description (whether caused by a pupil or question) occurs that cannot instantly be sorted out, then:

a) *check the time* and note it on the time sheet

b) stop the circus and get all pupils to the front of the room

c) solve the problem

d) restart the circus

If the problem occurred within the *first 45 seconds* of a new station, they restart at that station. If the problem occurred after 45 seconds, that station is ''lost'' for each pupil and you restart at the next move.

e) record the incident and outcome on the time sheet

Level of help

a) When requested, or when it is *obvious* (lack of activity, totally wrong activity) that a pupil is in difficulty with reading you may:

- — read out a specified word
- — read out the whole question word by word, and then at normal speed once

This does *not* apply to "key" questions in Category 3 (ie. those in sub-category 3α). Do *not* give reading help with them.

b) Very rarely, a pupil needs a bit of general encouragement with the first couple of stations to help him/her get going.

c) in Category 3 questions, pupils may be helped in the use of any apparatus — eg. how to heat a test tube correctly.

General

Controlling these circuses makes considerable demands upon the administrator. Throughout, it is essential that the administrator moves constantly around the circus. This does *not* distract the pupils and it greatly assists in keeping yourself "aware" of how the stations are progressing, and allows you to do some servicing before the 6 minutes are up if the pupils have finished. If it is necessary to talk to the resident teacher during the circus, keep an eye on the pupils at the same time.

Some problems that have arisen are — pupils who tag the whole pile of questions at a station instead of the one they have answered; pupils who take the eraser from the drawing question; pupils who use a watch for estimating time; pupils who use the whole 1 kg block of plasticine to estimate 20 g.

So you must be ready to deal with the unexpected!

Appendix 4

Administration and marking of written tests

Categories 1 (Using symbolic representations), 4 (Interpretation and application) and 5 (Design of investigations) were tested using pencil and paper tests.

Administration

These written tests and instructions on how to run them were sent to the schools by the National Foundation for Educational Research and administered by teachers in the schools.

Schools were asked to run the tests in the week 10th to 14th November 1980, allowing the next fortnight to follow up absentees.

Scripts, together with pupil data sheet and school questionnaire were posted back to the NFER, and sent up to Leeds after checking.

Each category had its own instruction page, following the front cover, to indicate to pupils the nature of the questions in the text.

The following is an example, the instruction page from Category 4 — Interpretation and application:

> The questions in this test are to find out how you think about problems. In some questions all the information you need is given to you. In others you need to use ideas you may have learned at home or at school.
>
> We are very interested in your ideas so try every question, even if you are not completely sure of the answer.

- Please read each question carefully.
- Put up your hand if you need help in reading a question.
- Each question is on a separate page. You answer some questions by putting a tick or cross in a box. In other questions you write your answer in the spaces provided.
- Work through the questions at your own rate — there will be enough time for you to try them all.
- We hope you will find the questions interesting.

Each test package was designed to last 50 – 60 minutes, but individual pupils varied greatly in the time they needed. Each test package was given to about 750 pupils.

In each test package questions were arranged in two different orders, one being the reverse of the other, to reduce any effects due to particular questions coming at the end of a test. In general, questions were grouped so that those with a similar rubric or mode of response occurred together.

Teachers administering the tests were asked to give help with reading when this was required — reading particular words, or reading the whole question once slowly, and then once at normal speed.

Marking

Marking meetings were held in January 1981 in three locations, Cambridge, Leeds and London.

At marking meetings, mark schemes were discussed, coding up of marks on computer sheets explained, and sample scripts marked and coded up. These were discussed, so that there was agreement on acceptable levels of inference and interpretation and any minor modifications to the mark schemes. Each marker had about 470 scripts, some of which had been premarked by the team, and these were checked on return of the coding sheets. Raw marks were entered on the coding sheets, all transformations being done by the computer. Some questions were also marked with a letter code (category marked) to indicate the nature of pupils' responses, particularly to try to highlight the types of errors being made.

Appendix 5

Notes on the sampling of schools and pupils

The selection of pupils for testing in this survey involved two stages of sampling. The first stage required the stratified random selection of a sample of schools with pupils in the 15 year old age group, the second the random selection of appropriate numbers of pupils from within each of these schools. In other words, the survey sample was selected by a two-stage stratified cluster sampling scheme.

At the school selection stage, the population of schools was first stratified by reference to four descriptive variables: region, location, school type and school size. In addition to Wales and Northern Ireland, there were three English regions (*North, Midlands* and *South*), defined as indicated at the beginning of Chapter 2. The location variable is relevant in England only, schools being classified into metropolitan or non-metropolitan locations depending on the county in which they were to be found (details of the county designations are also given in Chapter 2). The school type classifications used were: comprehensive to 16, comprehensive to 18, other maintained and independent. In addition, the comprehensive school sector was further subdivided into four size-of-age-group categories: 1 – 80, 81 – 160, 161 – 240 and 241 + .

Schools in England were selected from within each of the resulting sub-groups in numbers which reflected sub-group size. In other words, the proportion of "large Midlands comprehensives" in the survey sample was intended to reflect the proportion of such schools existing in the population as a whole. Wales and Northern Ireland were deliberately over-sampled to provide larger pupil sample sizes than would be the case with proportional random sampling of their schools.

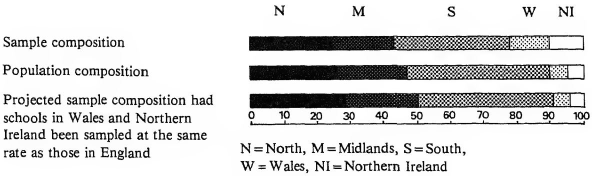
As Table A5.1 shows, 728 schools were actually invited to participate in the survey, and 568 of these eventually formed the participating sample.

Table A5.1 *School participation*

	England	Wales	Northern Ireland	All
Invited to take part	553	90	85	728
No reply	37	8	4	49
Unable to take part	71	10	20	101
Test not returned or returned unused	5	1	4	10
Number participating	440	71	57	568

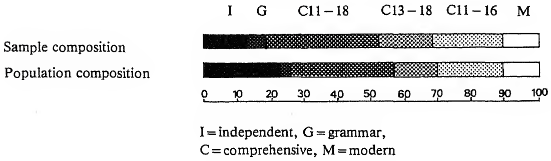
Figure A5.1 illustrates the regional composition of both the school sample and population.

Figure A5.1 *The regional composition of the school sample and population*



The compositions of the English school sample and population in terms of different types of school are shown in Figure A5.2. As explained in Chapter 2, for reporting purposes the comprehensive school sector was divided into three groups on the basis of age-range: 11 – 16 (including 13 – 16), 11 – 18 and 13 – 18. Figure A5.2 includes this distinction.

Figure A5.2 *Composition of English school sample and population in terms of type of school*



23 different test packages were administered in England and 21 in Wales and Northern Ireland. Each sample school took at least three different packages, and occasionally more than three if some of their pupils were to take two tests (there was a proportion of double testing of pupils in order to provide data for the later exploration of relationships between, for example, practical and written test performances).

The number of pupils in any school who would attempt the *same* test package was fixed at nine by reference to the size used for practical circuses. Survey schools in England were asked to provide 27 pupils in total for survey testing, schools in Wales and Northern Ireland being required to provide 36 pupils each.

Within each participating school, pupils were randomly selected on the basis of specific birth dates from all of those of the appropriate age in the school (ie. those born between 1 September 1964 and 31 August 1965 inclusive). To make the selection each school was asked to list all pupils of the relevant age who were born on the first day of any month, followed by all those born on the second day of any month, then those born on the third day, and so on until the required number of pupils was reached.

In Wales and Northern Ireland a number of the survey schools were also to take part simultaneously in the 1980 Language survey at age 15, and these schools were asked to select all pupils whose birthdays fell within a specified range of dates.

The names of the selected pupils remained known only to the school concerned all other parties having access only to the pupils' unique identifying numbers.

The only pupils explicitly excluded from the survey were those in special schools or in units designated as "special" within normal schools. However, the headteacher of each selected school was told that discretion could be used in withdrawing particular pupils from the testing sessions if it was felt that participation would cause undue distress.

One other procedural point which should be mentioned here is the method of handling pupil absentees. Pupils who were in the sample but were absent on the day on which the school undertook a written test session were given the relevant test if they attended school at any time within two weeks of the school's main test session. This was an attempt to avoid introducing an unnecessary bias into the pupil sample which would occur if, for example, persistent absentees happened also to be the lower performers in general. Economic reasons prevented such a procedure being adopted in the case of practical test sessions, which involved visits by trained administrators. For practical tests only, pupils absent on the day of a test session were replaced by reserve pupils who had previously been listed by the school.

16,301 pupils were tested in the survey, after the schools' withdrawal of 41 pupils and the absence of 848 others. 503 pupils were substituted for absentees in the practical test sessions. Clerical errors of one kind or another resulted in the overall loss of test results for about 3% of the pupils at the final analysis stage, the loss per test package varying between 1% and 5% (the most usual being 3% – 4%).

Table A5.2 provides details of the actual numbers of scripts which were involved at the analysis stage for each of the 24 survey test packages.

Table A5.2 *Pupil and school samples at the analysis stage*[1]

Sub-categories[4]	Test package	Pupils	Schools
1α, 1β	1	759	88
	2	739	90
	3	748	88
2α, 3β	1[2]	530	60
2α, 2γ	2[2]	536	62
2α, 3γ	3[2]	538	61
3β, 3γ	4	730	86
3β, 3γ	5	747	89
3β, 3γ	6[3]	200	25
4α	1	747	88
	2	728	87
	3	752	89
$4\delta_{AB}$	1	713	88
	2	752	90
	3	750	90
$4\delta_{CD}$	1	732	88
	2	754	85
	3	711	84
$4\delta_{EF}$	1	756	88
	2	763	88
	3	719	85
5α, 5β	1	737	89
	2	729	86
	3	738	86

[1] The numbers of pupils and schools sum to more than the actual total survey sample sizes, because schools took three or more of the test packages and many pupils within these schools attempted more than one test.

[2] These test packages were administered only in England.

[3] These test packages were administered only in Wales and Northern Ireland.

[4] Sub-categories 1γ, 2β, 2γ, 3α, 4β, 4γ, 4ε and 5γ were represented by individual questions contained in appropriate category test packages.

Appendix 6

SCHOOL NUMBER:

Assessment of Performance Unit, Department of Education and Science

ASSESSMENT OF PERFORMANCE IN SCIENCE

SCHOOL QUESTIONNAIRE

AGE 15 SURVEY

It is an essential part of the Science Monitoring exercise to relate the test results to information about the general provision for science in schools. This questionnaire is designed to gather such information.

If a question has a number of coded alternative responses none of which exactly describes your particular circumstances please select the one which is closest to the actual situation. If a particular question cannot be answered in the form required then please leave the appropriate box(es) blank and attach an explanatory note to the questionnaire.

Many of the questions require numerical answers which are to be entered into boxes provided. If your answer is, for example, a 3-figure number which is to be entered into a 4-column box, then use the last three columns of the box leaving the first column blank.

We realise that completing the questionnaire is an extra burden imposed on you, but we hope that you will understand that it is only with such information that full use can be made of the survey results.

© University of Leeds

© Chelsea College, University of London
0326/0328.

APU SCIENCE: AGE 15 SURVEY 1980

SCHOOL QUESTIONNAIRE

---

1. Please provide the following general school information:
   (a) *What is the age range of your school?*
   (b) *How many pupils are on your school roll?*

1.

a				b			
8	9	10	11	12	13	14	15

TEACHERS AND SCIENCE-TEACHING RESOURCES

2. For the purpose of this question, a formally qualified science teacher is defined as a science or applied science graduate *or* a teacher with a teaching certificate with a main course in a science subject.
   (a) *How many "full time equivalent" teachers do you have on your staffing establishment?*
   (b) *How many of these teachers are currently involved in teaching science?*
   (c) *How many of the teachers in (b) hold formal science qualifications?*

2.

a			b		c	
16	17	18	19	20	21	22

3. For the purpose of this question, a laboratory is to be considered as a working space where pupils can do small-group and individual practical work with services (gas, water, electricity).
   (a) *How many laboratories does your school have?*
   (b) *How many of these laboratories can accommodate 20 or more pupils?*
   (c) *How many technicians work in your science department?*

3.

a		b		c
23	24	25	26	27

4. When answering this question, please count a double teaching period as two single periods.
   (a) *How long, in minutes, is a single teaching period in your school?*
   (b) *How many periods are there in your timetable cycle?*

4.

a		b	
28	29	30	31

5. Do any of your staff use one or more of the following out-of-school resources as part of science courses up to age 16? Use codes Y or N as appropriate in each case.
   (a) *Field centres*
   (b) *Industrial visits*
   (c) *Link courses with colleges of F E*
   Other — please specify:

5.

a	b	c
32	33	34

6. For each of the technicians indicated in 3(c), please provide the following information (use one set of boxes for each technician):
   - (a) *Total hours/week the technician is employed*
   - (b) *Total weeks/year the technician is employed*
   - (c) *Does the technician work during school holidays?* Use codes Y or N.
   - (d) *Please estimate the hours/week spent by the technician in work for non-science departments*

6.

a		b		c	d	
35	36	37	38	39	40	41
42	43	44	45	46	47	48
49	50	51	52	53	54	55
56	57	58	59	60	61	62
63	64	65	66	67	68	69
70	71	72	73	74	75	76

FINANCIAL RESOURCES

If yours is an independent school, please provide the total amount of money received by your school for teaching resources as an answer to question 7 and ignore question 9.

Financial Years

7. Please indicate the total amount spent by the school in each year from the LEA allowance on teaching resources.

7.

1977/78					1978/79					1979/80				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

8. What was the amount spent by the science (or science subject) departments from this allowance?

8.

77/78				78/79				79/80			
16	17	18	19	20	21	22	23	24	25	26	27

9. If any additional grant was received by the science department from any source (eg. LEA, industry, parents' association) please indicate the total amount here.

9.

77/78				78/79				79/80			
28	29	30	31	32	33	34	35	36	37	38	39

10. Please give a rough indication of how the total science department allowance is apportioned between the following three kinds of expenditure:
   - (i) Paper, Stationery, and consumable reprographic materials
   - (ii) Text books and resource books (excluding those in school library)
   - (iii) Apparatus, equipment and consumable materials other than those in (i)

10.

	77/78				78/79				79/80			
i	40	41	42	43	44	45	46	47	48	49	50	51
ii	52	53	54	55	56	57	58	59	60	61	62	63
iii	64	65	66	67	68	69	70	71	72	73	74	75

11. If your school has been involved in any kind of restructuring (eg. a merger with another school) which would explain major changes in your budget figures over the last three years, please describe the circumstances:

---

12. For schools in Wales only:
    (a) *Please use the appropriate bracketed code letter to indicate the medium (spoken and written) used by pupils in science in your school — entirely Welsh* (W), *entirely English* (E), *both Welsh and English* (B).
    (b) *If the science teaching medium is a combination of Welsh and English, then please further indicate the approximate percentage of work done in Welsh.*

12.

a	b	
77	78	79

COMMENTS

If you would like to comment on any difficulties you experienced in completing the questionnaire, or if you would like to provide information not asked for in the questionnaire but which you feel relevant, then please do so here:

Appendix 7

Notes on the survey analysis

The survey data is held on the Amdahl V/7 computer at the University of Leeds and all analyses were conducted on this machine.

Question detail

The individual question results reported in Chapters 3 to 8 are raw sample statistics and were produced with the aid of SPSS (Nie *et al,* 1975). The function of this information in the report is principally to illustrate the nature and variety of tasks embraced by the titles of the various science sub-categories.

The school questionnaire

The school questionnaire data was also analysed using SPSS. As the school sample was produced by a simple random sampling procedure within each stratum it would be appropriate to apply the usual chi-square and other significance tests to the data if the statistical significance of any of the sub-group differences is of interest.

The sub-category mean performance estimates

Population mean performance estimates and their associated estimated variances were produced test package by test package for the sub-categories contained in them. Where a sub-category was represented in two or more different test packages, the separate, independent performance estimates produced from these were combined.

When producing the population estimates for each science sub-category in each test package (after initially adjusting all question scores onto a common mark scale — as explained in Appendix 2) it was necessary to weight the raw sample data to take account of the stratified cluster-sampling scheme which was used in sample selection. The method adopted here to produce a population mean estimate for any particular sub-group ("North", comprehensive 11–16, etc) was first to multiply each school mean by that school's size (ie. by the number of pupils in that school of the appropriate age), to sum the resulting figures over all schools in that sub-group sample, and then to divide the whole by the total number of pupils of the appropriate age in those sample schools (technically termed a biased ratio estimate). Estimates were produced in this way for sample breakdowns by the stratifying variables, ie. school location, school type and region. Weighted variance estimates were produced simultaneously — the relevant formula being found in any text on survey sampling (for example, Kish 1965, Cochran 1977, ch.9; Hansen, Hurwitz and Madow 1953, Vol. 2, ch.6; Deming 1960).

The overall population mean and variance estimates, and those for boys and girls separately, were produced by appropriately weighting the separate regional estimates before combining.

The sub-category mean estimates reported in Chapter 9 were produced by averaging the appropriate sub-test estimates, and the standard errors associated with these final mean scores were produced in the usual way for an average of independent variates by dividing the square root of the sum of the separate variance estimates by n (where n is the number of sub-tests involved).

Design effect

Kish (1965) first introduced the notion of the *design effect* of a survey sampling scheme. The design effect of a sampling scheme indicates the degree to which population estimates based on the resulting sample are more or less accurate than they would be had they been based on a simple random sample of similar overall size.

A stratified sampling scheme, for example, can often result in a more efficient sample in estimation terms than a simple random sampling procedure, while a cluster sampling scheme usually produces a less efficient sample in this sense. The design effect is defined as the ratio of the variance associated with a population estimate based on the sample produced by a particular sampling scheme to the variance which would have been associated with that estimate had it been based on a simple random sample of similar size.

The design effects for this survey vary from sub-category to sub-category, ranging from 2.0 to 3.6. This means that the standard errors quoted in Chapter 9 are up to about twice as large, and consequently the confidence intervals up to twice as wide, as they would have been had the performance estimates concerned been based on similarly-sized simple random samples of pupils.

Raw sample standard errors were doubled to produce estimated standard errors for the performance estimates of pupil sub-groups for which no population information was available (eg., for pupils attending schools with an average science expenditure of £1 per pupil per year). In other words, the standard errors produced by treating the respective pupil sub-samples as though they were simple random samples were doubled in these cases.

Links between sub-category results and questionnaire data

Data resulting from a sample survey of this kind, involving the administration of many different test packages, has limited potential for the exploration of relationships between test performances and various pupil and school characteristics. In this particular survey the potential for this kind of work was rendered rather more limited by the need to incorporate a pilot survey element into the package administration design. Each test package was administered to about nine pupils in any school, and the total school sample for any one package was usually between 60 and 90 schools. Just over half of the schools in each of these sub-samples would be comprehensive schools in England. It will be appreciated that with these small numbers there were often insufficient numbers of schools in every cross-classification cell to allow any investigation of association between test performance and some of the school background variables. Also it was noted in Chapter 2 that there was often little variation across schools for certain of the potential background variables. In view of this, it should be said that the absence in this report of evidence of associations of one kind or another does *not* imply that no such association in fact exists. Where the statistical significance of a comparison is indicated, this is at the 5% level.

Table A7.1. *Summary of some of the survey results*

			Sex		*Region*[1]					*Location*[2]		*School size*[3]			*Pupil-teacher ratio*[3]			*Percentage uptake of free school meals*		
		Overall	Boys	Girls	North	Midl	South	Wales	NI	met	non-met	< 600	601 – 1,000	1,001 +	1 – 14.9	15 – 17.4	17.5 +	< 3%	4% – 6%	> 6%
1α	mean	63.3	65.5	61.4	60.2	65.3	65.4	58.5	61.6	61.8	64.6	57.5	63.9	63.2	59.7	64.0	62.2	66.4	64.6	58.5
	s.e.	0.9	1.3	1.2	1.7	1.7	1.7	1.4	1.7	1.7	1.3	3.5	1.7	2.0	2.6	1.6	2.4	2.0	2.4	1.9
1β	mean	56.1	57.7	54.8	53.0	59.3	57.9	49.7	51.1	54.8	57.4	53.1	56.3	56.0	51.6	57.0	56.2	60.2	58.5	50.1
	s.e.	1.1	1.6	1.6	2.0	2.3	2.0	1.7	2.1	2.0	1.6	4.3	2.2	2.4	3.6	2.0	3.0	2.5	3.0	2.5
2α	mean	40.6	43.2	37.5	39.6	39.0	42.1	–	–	38.8	41.5	–	39.1	40.8	38.4	39.6	40.8	44.0	38.4	37.3
	s.e.	0.9	1.3	1.3	1.3	2.0	1.5			1.4	1.2		1.6	1.6	2.4	1.6	2.0	2.0	2.4	1.6
3β	mean	40.0	39.9	40.0	40.1	39.3	41.7	36.5	33.8	37.9	42.2	–	38.2	42.1	40.1	38.9	42.1	39.1	36.9	33.4
	s.e.	0.8	1.1	1.2	1.5	1.4	1.4	1.8	1.8	1.2	1.1		1.3	1.5	2.1	1.5	1.6	2.0	2.2	1.8
4α	mean	44.3	46.0	42.6	44.5	42.1	46.1	41.2	40.4	43.7	45.5	46.9	45.0	44.1	44.2	44.3	45.1	47.0	45.7	41.1
	s.e.	0.8	1.2	1.1	1.3	1.9	1.3	1.1	1.5	1.2	1.3	5.1	1.6	1.6	2.5	1.6	2.1	2.0	2.1	1.7
$4\delta_{AB}$	mean	41.6	42.4	40.8	40.3	44.9	41.3	39.0	39.7	40.7	42.4	39.9	40.4	42.1	37.9	41.6	41.9	43.7	43.7	37.6
	s.e.	0.8	1.1	1.2	1.6	1.3	1.3	1.1	1.3	1.5	1.0	3.4	1.4	1.7	2.5	1.4	2.0	1.8	2.2	1.7
$4\delta_{CD}$	mean	35.1	39.8	30.2	34.0	36.4	36.0	31.9	31.6	33.7	36.3	33.7	34.4	35.5	32.6	34.4	35.8	37.0	36.1	30.9
	s.e.	0.7	1.1	0.9	1.3	2.1	1.0	1.1	1.3	1.2	1.0	3.5	1.4	1.4	2.3	1.3	1.8	1.6	2.3	1.4
$4\delta_{EF}$	mean	38.4	39.4	37.2	37.9	40.8	38.2	35.1	35.3	36.5	40.0	36.8	38.0	39.1	32.1	39.1	40.2	40.6	40.0	34.8
	s.e.	0.9	1.3	1.2	1.8	1.7	1.5	1.3	1.6	1.5	1.2	4.0	1.7	1.6	2.7	1.5	2.0	1.9	2.5	1.8
5α	mean	61.4	61.4	61.0	59.9	61.3	63.6	57.2	55.9	59.6	63.2	60.1	61.0	61.4	59.5	61.2	62.5	64.0	62.0	57.3
	s.e.	0.9	1.3	1.3	1.5	1.6	1.7	1.6	2.0	1.7	1.1	4.0	2.2	2.2	5.0	1.7	2.9	2.1	3.0	2.0
5β	mean	48.7	48.5	48.6	46.6	48.3	50.9	46.7	45.0	46.8	50.0	48.2	47.5	48.9	45.5	48.7	48.6	50.6	49.2	44.4
	s.e.	0.8	1.1	1.0	1.5	1.5	1.6	1.3	1.7	1.4	1.2	3.5	2.0	1.9	4.0	1.5	2.4	1.8	2.5	2.0

[1] Category 2 testing not undertaken in Wales or Northern Ireland

[2] England only

[3] Comprehensive schools in England: insufficient sample for practical category estimations for smallest school size

Appendix 8

Standard errors, confidence limits and statistical significance

The mean sub-category performance figures presented in Chapter 9 are based on a sample of pupils, and therefore are estimates of the mean performance levels of the population of pupils in the various sub-groups discussed in that Chapter. The accuracy of a particular estimate will depend on the size of the pupil sample on which it is based, on the degree of "clustering" of pupils in schools allowed in the sampling scheme (ie. on the number of pupils taking the same test package in each school) and on the degree of variation of performance found among these pupils and schools.

The standard error associated with a particular performance estimate provides an indication of its accuracy. On the basis of statistical theory, there is a 95% probability that the true mean performance level will be within a range of about two standard errors either side of its estimated value. Thus, if we have a mean performance estimate of 61% with an associated standard error of 1.6, then we can feel 95% certain that the actual population level of mean performance for that pupil sub-group lies within the range 61± 3.2, ie. within the range 57.8 to 64.2. This range is called the 95% confidence interval for the performance estimate, and these intervals have been shown in the graphical presentations of Chapter 9.

A difference in the estimated mean performance levels of two pupil sub-groups is said to be statistically significant at the 5% level if such a difference could in theory occur by chance at most five times in a hundred. In order to establish whether or not a difference in estimated mean performance levels is statistically significant, we must first calculate the standard error associated with that difference. If a performance difference is larger than twice its standard error, then it is said to be statistically significant (at the 5% level). A useful visual indicator of statistical significance is provided by the confidence intervals of the performance estimates. Where the confidence intervals do not overlap, or overlap only very slightly, then the performance estimates concerned are likely to be significantly different from each other in the statistical sense.

A cautionary note to readers might be appropriate at this point.

Firstly, the criterial level of statistical significance applied throughout this report is a fairly arbitrary, though commonly used, one. If a more severe one had been adopted (the 1% level, for example) then many of the performance differences identified as statistically significant in this report could lose this "status". Similarly, had a more lenient level been applied (the 10% level, say) then many performance differences which were not commented upon would become statistically significant.

Secondly, remaining with the 5% level of statistical significance, many of the performance differences which proved not to be statistically significant had high standard errors associated with them. It could be that population differences in performance levels actually exist in some cases, but that the sample sizes available for performance estimation were too small for these differences to be detected. Sometimes, for example, despite the fact that few individual performance differences reach statistical significance, the consistency in the direction of performance differences across the profile of sub-categories can be taken to suggest an underlying population difference.

Finally, it should be recognised that statistical significance is not synonymous with educational significance, and this has a bearing both on the interpretability of many of the findings and on their implications for science education.

Appendix 9

Administrators of practical tests

Mr G Ashford	Royton & Crompton School, Royton, Oldham.
Mr R S Barrett	The Meads School, Chertsey, Surrey.
Mr M Baxter	Advisory Teacher for Science, Northamptonshire Education Dept.
Mr A Bevan	Beverley School, New Malden, Surrey.
Mr P R Bishop	Kentwood School for Boys, Penge, Kent.
Mr R G Briggs	Down High School, Downpatrick, Co. Down, Northern Ireland.
Mr D Callear	Alexandra Park School, London N22.
Mr N Catanach	North Chadderton School, Oldham, Lancs.
Dr S Chinn	Kings of Wessex Upper School, Cheddar, Somerset.
Mr M B Clemmet	Cordaux High School, Louth, Lincoln.
Mr E Cross	Walton-le-Dale High School, Bamber Bridge, Preston, Lancs.
Mr G Davies	Ysgol Brynrefail, Llanrug, Gwynedd, Wales.
Mr R Edwards	Brockworth School, Brockworth, Gloucestershire.
Mr B T Foster	Buckler's Mead School, Yeovil, Somerset.
Mr M Frost	Bishop Stortford Boys' High School, Bishop Stortford, Herts.
Mr T Greenall	All Hallows R.C. School, Penwortham, Preston, Lancs.
Mr A C Hall	Pocklington Woldgate School, Pocklington, York.
Mrs I Hall	Cross Hall High School, Ormskirk, Lancs.
Mrs J Hall	High Cross School, Tottenham, London.
Mr M R Hall	Whickham Comprehensive School, Whickham, Newcastle-upon-Tyne.
Mr W Hart	Kingsbury High School, Kingsbury, Tamworth, Staffs.
Mr K Holmes	Brayton High School, Selby, North Yorkshire.
Mr K Holt	Creighton School, London N10.
Mr B Jones	Tonyrefail Comprehensive School, Tonyrefail, Mid Glamorgan.
Mrs H Jones	Tolworth Girls' School, Tolworth, Surbiton, Surrey.
Mr J I Jones	Ysgol Tregaron, Tregaron, Dyfed, Wales.
Mrs B Livesey	Witton Park High School, Blackburn.
Mr D J Marjoram	St Patrick's College, Armagh, Co. Armagh, Northern Ireland.
Mr B Mason	Wolgarston High School, Penkridge, Staffs.
Miss R McCall	Killey and Stone School, Newark, Notts.
Mr G J Monney	Laindon School, Laindon, Basildon, Essex.
Mr A Nicholl	Coleraine Academical Institution, Coleraine, Co. Londonderry, Northern Ireland.
Mr J Oversby	Stokesley Comprehensive School, Stokesley, North Yorkshire.
Mr C Proctor	Evesham High School, Evesham.
Mr S Quail	Bullers Wood School for Girls, Chislehurst, Kent.
Mr B J Seabourne	Henry Beaufort School, Harestock, Winchester, Hants.
Miss C Voller	Coombe Girls' School, New Malden, Surrey.
Mr R Whitehouse	The Vale of Ancholme School, Brigg, Humberside.
Mr J Wilcox	Smestow Comprehensive School, Castlecroft, Wolverhampton.
Mr M Williams	Barstable School, Basildon, Essex.

Appendix 10

APU Steering Group on Science

Mr T A Burdett, HMI	Chairman
Professor P J Black	Director, Centre for Science Education, Chelsea College, University of London
Mr N B Evans, HMI	Welsh Education Department, Cardiff
Mr A R Hall (to Easter 1982)	Sir Joseph Williamson's School, Rochester
Sister M Hurst (to August 1982)	Schools Council
Mr J R Jeffery (from November 1981)	Pocklington School
Professor R F Kempa (from September 1982)	University of Keele
Dr W J Kirkham (from Easter 1982)	Leicester LEA
Professor D Layton (to May 1982) (Professor D Child – Director from May 1982)	Director, Centre for Studies in Science Education, University of Leeds
Mr I Milligan, HMI	Department of Education, Northern Ireland
Mr C Parsons, HMI	HM Inspectorate
Dr B M Prestt	Manchester Polytechnic
Mr J C Taylor (to Easter 1982)	Manchester LEA
Mr H D Wilcock (from Easter 1982)	Paganel Junior School, Birmingham
Mr K Wild (to Easter 1982)	Staffordshire LEA
Mr G D Williams (to Easter 1982)	Wellesbourne County Primary, Liverpool

Appendix 11

Monitoring Services Unit at the time of the Age 15 science survey 1980

Mrs Barbara Bloomfield	Head of Unit
Mrs Anne Baker	Deputy
Miss Susan Darby	
Mrs Mary Hall	
Miss Elizabeth Evans	Secretary

Appendix 12

Members of the APU Consultative Committee

Professsor J Dancy (Chairman)	School of Education, University of Exeter
Miss J E L Baird	Joint General Secretary, AMMA
Mr P Boulter	Director of Education, Cumbria (ACC)
Mrs J Bushy	Council, Bromsgrove District Council
Mr P J Casey	Deputy Director (Education and Training), CBI
Mr R G Cave	Former Senior Education Officer, Cambridgeshire
Mr L Cooper	Deputy Headmaster, Sherburn High School (NAS/UWT)
Mr H Dowson	Deputy Headmaster, Earl Marshal School, Sheffield (NUT)
Professor S J Eggleston	Department of Education, University of Keele
Mr P J Eley	National Confederation of Parent-Teacher Associations
Mr A Evans	Education Deparment, NUT
Mr D Fisher	County Education Officer, Hertfordshire (ACC)
Mr G S Foster	Headmaster, The Towers School, Ashford (NUT)
Mr G Hainsworth	Director of Education, Gateshead (AMA)
Mr K S Hopkins	Deputy Director of Education, Mid-Glamorgan (WJEC)
Councillor P Horton	Sheffield Metropolitan District Council (AMA)
Mr C Humphrey	Director of Education, Solihull (AMA)
Dr K Jones	Parent and Doctor, Sheffield
Mr T M Jones	Headmaster, Werneth Junior School, Oldham (NUT)
Mr J A Lawton	Kent County Council (ACC)
Mr G M Lee	Doncaster Metropolitan Institute of Higher Education (NATFHE)
Mr S Maclure	Editor, Times Educational Supplement
Mrs R Mills	Consultant Economist
Mr M J Pipes	Headmaster, City of Portsmouth School for Boys (NAHT)
Dr W Roy	Headmaster, The Hewett School, Norwich (NUT)
Professor M D Shipman	Department of Education, University of Warwick
Miss A C Shrubsole	Principal, Homerton College
Mr F A Smithies	Assistant General Secretary (Education), NAS/UWT
Mr T P Snape	Headmaster, King Edward VI School, Totnes (SHA)
Professor J Wrigley	School of Education, University of Reading
Mr A Yates	Director, National Foundation for Educational Research

Appendix 13

Note on the APU

The Assessment of Performance Unit (APU) was set up in 1975 within the Department of Education and Science. It aims to provide information about general levels of performance of children and young people at school and how these change over the years.

The terms of reference of the APU are:

> "to promote the development of methods of assessing and monitoring the achievement of children at school, and to seek to identify the incidence of under-achievement".

Associated with these terms of reference are the following tasks:

1. To identify and appraise existing instruments and methods of assessment which may be relevant for these purposes.
2. To sponsor the creation of new instruments and techniques for assessment having due regard to statistical and sampling methods.
3. To promote the conduct of assessment in co-operation with local education authorities and teachers.
4. To identify significant differences of achievement related to the circumstances in which children learn, including the incidence of under-achievement, and to make the findings available to those concerned with resource allocation within government departments, local education authorities and schools.

The APU monitoring programme has been concerned to reflect the breadth of the curriculum in schools and to display the wide range of pupil performance. The assessment model adopted by the APU is based on a number of curricular areas. At the present time, monitoring is proceeding in mathematics, language and science, and the first survey of children's performance in their first foreign modern language will take place in 1983. The assessment of pupils' abilities in the field of design and technology is under consideration, but there is no commitment to monitoring in this area.

The assessment procedures in science are developed jointly by the Science Monitoring Teams at Chelsea College, London and the University of Leeds. The work is steered by a group consisting of teachers, advisers, teacher trainers, educational researchers and HMI. The work of all groups and the progress of the monitoring surveys is supervised by a small management team, which includes the administrative and professional heads of the Unit together with the chairman of each group. More generally, the APU is advised by a Consultative Committee, which is broadly representative of education, industry, commerce and parental concern.

In producing national pictures of aspects of pupil performance, the Unit does not report on the performance of individual pupils, schools or local eduction authorities. Survey results are published regularly, in the form of reports, by Her Majesty's Stationery Office. This is the first report to be published on the science performance of 15 year olds. Reports on the first surveys of science performance at ages 11 and 13 have already been published.

Further information about the work of the APU is available from the Department of Education and Science, Information Division, Room 211, Elizabeth House, York Road, London EC1 7PH.

Printed in England for Her Majesty's Stationery Office
by Hobbs the Printers of Southampton
(2205) Dd736055 C30 12/82 G380